

Clinical Decision Support Systems (CDSS) Applications in Psychological Suitability Assessments

by

Kevin M. Kennedy

Thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Arts (MA) in Psychology

The Faculty of Graduate Studies
Laurentian University
Sudbury, Ontario, Canada

© Kevin Kennedy, 2017

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE

Laurentian University/Université Laurentienne
Faculty of Graduate Studies/Faculté des études supérieures

Title of Thesis
Titre de la thèse CLINICAL DECISION SUPPORT SYSTEMS (CDSS) APPLICATIONS IN
PSYCHOLOGICAL SUITABILITY ASSESSMENTS

Name of Candidate
Nom du candidat Kennedy, Kevin

Degree
Diplôme Master of Arts

Department/Program
Département/Programme Psychology Date of Defence
Date de la soutenance September 11, 2017

APPROVED/APPROUVÉ

Thesis Examiners/Examineurs de thèse:

Dr. Cynthia Whissell
(Supervisor/Directeur(trice) de thèse)

Dr. Michel Larivière
(Committee member/Membre du comité)

Dr. Normand Gingras
(Committee member/Membre du comité)

Dr. Peter Hausdorf
(External Examiner/Examineur externe)

Approved for the Faculty of Graduate Studies
Approuvé pour la Faculté des études supérieures
Dr. David Lesbarrères
Monsieur David Lesbarrères
Dean, Faculty of Graduate Studies
Doyen, Faculté des études supérieures

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

I, **Kevin Kennedy**, hereby grant to Laurentian University and/or its agents the non-exclusive license to archive and make accessible my thesis, dissertation, or project report in whole or in part in all forms of media, now or for the duration of my copyright ownership. I retain all other ownership rights to the copyright of the thesis, dissertation or project report. I also reserve the right to use in future works (such as articles or books) all or part of this thesis, dissertation, or project report. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that this copy is being made available in this form by the authority of the copyright owner solely for the purpose of private study and research and may not be copied or reproduced except as permitted by the copyright laws without written authority from the copyright owner.

Abstract

Psychological suitability assessments are an integral component of public safety recruitment and selection processes. Psychological suitability assessments can benefit from the implementation of decision support systems. Providing an estimate of the likelihood for outcomes can act to support psychologists in decision making processes. This thesis critiques a content focused approach to psychological suitability assessment, develops a classification algorithm that estimates profile outcome likelihoods of several possible psychological suitability decision categories, and assesses the effect of providing estimates to psychologists on psychological decision making processes when determining profile outcomes. The critique outlines that the assessment model dimensions were developed to suit the needs of the organization at the time and connected psychological data with anticipated behaviours and activities. However, the model was not developed within a systematic process and relied heavily on subjective accounts of on duty officers. The assessment model could be improved with the implementation of new and updated tools, such as the MMPI-2RF or the MPULSE. The classification algorithm established by the discriminant function analysis accurately classified 86.75% of cases ($p < 0.0001$) in this assessment model. In testing the effect of providing outcome estimates to psychologists to assist in their decision making processes, a MANOVA analysis was conducted. When ($n=14$) psychologists were presented with a likelihood estimate that was in support of the expert decision, psychologists significantly improved in their overall match accuracy ($p > 0.05$). Experience conducting these types of assessments was a significant covariate for the time it took to render a decision, and the time it took to render a decision was not affected by the presentation of a likelihood estimate. Overall, this study suggests that decision support systems can be implemented to support a temporary clinical lead in navigating the psychological data such that candidates are evaluated more consistently for psychological suitability

evaluations. Implementing a decision support system in practice can act as a guide for interpreting psychological data, reduce the error of both experienced and inexperienced assessors, and improve the integrity of the assessment.

Acknowledgements

First and foremost, I would like to dedicate my work to the late Frank Kennedy. He taught me that success isn't measured by what you do, but who you are. To this day, he continues to inspire me to fill my heart with gratitude and to work on something I love every day. Second, through all the changes and challenges our family has faced I would like to thank my mother, Lee-Anne, for sharing her seemingly endless love and joy that I have been fortunate enough to carry with me through thick and thin.

I am eternally grateful for Dr. Normand Gingras as a friend, mentor, and committee member. He has been the rutter to my voyage. Normand reminded me that in occasions when the sea is rough, you can easily lose sight of the destination. However, just because you can't see the destination, it doesn't mean that you are lost. He taught me that it is important to have courage, keep your eyes on the task at hand, and to keep moving forward even if you don't know where you'll end up. I would like to thank Dr. Whissell, who as a supervisor provided crucial support and freedom to explore my topic of interest. With her support, I could balance my life in my own image, while completing my work at a distance. I will always be grateful for her flexibility, compassion, and detailed review. She has helped me improve my understanding of science as a discipline and guided me on my path in the research community. I am also grateful to Dr. Lariviere for making the time to follow the work of a developing student, investing both his time and effort to provide a thorough review from his expertise on the psychological aspects of work in the world of federal correctional officers.

I am thankful to Dr. Laura Goodine, Sebastien Girard, Philip Yelle, and everyone else at the PSC for the times we have had together as colleagues and friends, as well as their overwhelming encouragement to pursue a Master's degree that pushed me to apply at Laurentian.

Dr. Katie McCann for her friendship during my time living in Toronto and her valuable support as a voice actor for my experimental brief, I'll always remember our extensive discussions on both medicine and psychology, and wish her the best in her career in gynaecology. I am sure she will make a difference in many others' lives, as she has in mine.

In addition to my personal benefactors, I would not have been able to complete my work on time if it wasn't for Mr. Donald Harvey who provided assistance integrating my python program into an online platform and Dr. Denis Lapalme for the time reviewing and discussing his content focused 6-dimension model used to assess public safety officers.

I feel privileged to have had a wealth of love and support from my family, committee members, friends, and colleagues throughout this process. I will always be grateful to everyone who has helped me on this journey.

Table of Contents

General Introduction	1
Chapter 1 - Six Dimension Model of Psychological Suitability	4
Introduction	4
Canadian Federal Public Safety.....	6
Six Dimension Assessment Model.....	7
Assessment Procedure	7
SIOP Principles	9
Assessment Model.....	9
Development.....	13
Validation	19
Conclusion.....	21
Strengths	21
Weaknesses.....	22
Chapter 2 - Classification Algorithm: Discriminant Function Analysis.....	24
Introduction	24
Procedure.....	25
Population.....	25
Data Cleaning	26
Data.....	26
Outcomes	26
Statistical Analysis	27
Results	28
Canonical Functions	28
Classification	32
Conclusion.....	33
Chapter 3 - Clinical Decision Support Systems for Psychological Suitability Assessment ...	35
Introduction	35
Theoretical Background	36

Clinical Decision Support Systems	36
CDSS for Psychological Suitability	37
Selecting profiles for the CDSS evaluation.....	38
Methods	40
Securities, Data Storage and Data Export	40
Instructions, Consent, and Experience Survey	41
Participant Population	42
Procedure	42
Results	49
Limitations	52
Conclusion.....	54
Discussion	56

General Introduction

This thesis is divided into three supporting chapters to investigate the utility of clinical decision support systems applied to psychological suitability assessments for public safety officers. Public safety officer candidates are required to meet psychological criteria to be employed in most public safety institutions in Canada. Personality is just one component, which is assessed when candidates apply for these types of positions; psychological data is integrated with other hiring information before a final decision is made. Psychological suitability evaluations are administered and interpreted by registered psychologists and/or psychological associates with training in psychometric theory as it is applied to public safety suitability assessment. The most documented and supported tool for assessing public safety candidates in the United States of America is a measure of psychopathology, the Minnesota Multiphasic Personality Inventory 2 (MMPI-2) (Aamodt, 2004). The MMPI-2 consists of 567 True-False items, providing the content for clinical scales measuring various types of psychopathology. Canadian public safety organizations apply a wide range of techniques and instrumentation to assess the probable behaviours, thoughts and emotions of public safety officer candidates, including the MMPI-2.

The first chapter within the thesis outlines and analyses a content-focused six dimension model of assessing psychological suitability. It opens a discussion on sources of variance in suitability recommendations. It is especially important in federal selection processes for public safety positions that candidates are assessed consistently. The six dimension outcomes, requirements and development processes were outlined by Dr. Denis Lapalme in a series of interviews he conducted with OPP applicants over several years. The process of development and validation of the system is evaluated against the standards set forth by the Society of

Industrial and Organizational Psychology's principles for psychological suitability assessment processes.

In the six dimension model, candidate psychological profiles are initially evaluated to determine whether the candidate meets the requirements, requires further review in an interview, or is unsuitable for public safety work. The second chapter builds on the first; using the PSC data collected applying this assessment model in the psychological suitability determination of 1800 correctional officer candidates from 2012-2015 for a discriminant function analysis to develop a classification algorithm for the initial evaluation of the profile. The chapter determines the relationships between key scales of the MMPI-2 and seven outcomes, then applies the classification algorithm back to the data and assesses the accuracy of predicting an expert decision from MMPI-2 psychological data. A quadratic discriminant analysis was completed to determine if primary profile decisions could be predicted from MMPI-2 profile data and to develop a method of estimating a profile's most likely primary decision in the context of the six dimension model. All the scales present in the dataset provided by the Public Service Commission were included in the analysis to allow for the highest degree of predictive accuracy of the linear combination.

The classification algorithm was implemented into an online experiment, where fourteen psychologists from across Canada made psychological suitability decisions for a sample of profiles either with or without information from the algorithm. The third and final chapter describes the experimental process, controls, and results of implementing outcome probability estimates as clinical decision support systems in a psychological suitability context. The experiment determined that clinical decision support systems applied to psychological suitability assessments significantly affects assessor decisions. Psychologists used the additional

information provided as decision support, but did not completely rely on it, making their own judgement of the outcome for each profile. It was determined that decision support systems can successfully be implemented to support psychologists in navigating the psychological data such that candidates are evaluated consistently for psychological suitability evaluations.

Chapter 1: Six Dimension Model of Psychological Suitability

Introduction

This chapter outlines and critiques an existing model that applies a six dimension framework to assess public safety candidates for psychological suitability. The chapter outlines the purpose of public safety psychological suitability evaluations, their role in selection processes, common instruments used, and Canadian federal public safety organizations that apply this technique in the selection of their front line staff. The six dimension content focused model is outlined from a series of interviews with its developer, Dr. Denis Lapalme. The assessment model's development and validation procedures are contrasted with the Society of Industrial and Organizational Psychology's guidelines for psychological suitability assessment procedures. Finally, this chapter concludes with an overview of the critique and practical implications for improving the six dimension model.

Public safety candidates are evaluated on personality traits, behaviour patterns, and psychological characteristics that might be problematic in the workplace. Personality data is often integrated into an evaluation of a candidate's overall suitability for a position. A suitable candidate, in public safety pre-employment assessments, is understood to demonstrate "personal factors that effectively contribute to human reliability in the particular occupation and the absence of factors that might pose a risk to the individual or to society" (CPA, 2013, p. 2). While candidates identifying a presence or history of mental illness may meet job selection criteria in personality assessments, psychological suitability evaluations are conducted to identify behavioural patterns that are potentially problematic which may or may not be present in various diagnostic categories. Candidates are evaluated to determine if their behavioural patterns are

likely to meet the minimum requirements to perform the responsibilities of the position (CPA, 2013). Psychological suitability evaluations are administered and interpreted by registered psychologists and/or psychological associates with training in psychometric theory and practice applied to public safety positions. The most documented and supported tool for assessing public safety candidates in the United States of America is a measure of psychopathology, the Minnesota Multiphasic Personality Inventory 2 (MMPI-2) (Aamodt, 2004). The original MMPI was developed in 1937 by Starke R. Hathaway and J. Charnley McKinley in the United States as an “objective aid in the routine psychiatric case work-up of adult patients as a method of determining the severity of the conditions” (Dahlstrom et. al., 1972). The MMPI became widely accepted as the standard objective measure of personality and psychopathology (Dahstrom, 1992). Over time, practitioners and researchers supported a revised version of the tool for national norms representative of the U.S. population with updated item content and reductions in clinical scale correlations (MMPI History, 2015). A committee was established to build what would become the MMPI-2, publishing the re-standardized MMPI, the MMPI-2, in 1989 (Butcher et al., 1989). The MMPI-2 has been used as a method of identifying diagnostic classification and rating scales in psychiatry (Butcher et al., 1989) as well as a personality test for employment of emergency personnel, especially when employees are required to carry weapons and mental stability and fitness can be reasonably related and necessary to the job (Principles, 2003). The MMPI-2 consists of 567 True-False items, providing the content for clinical scales measuring various types of psychopathology. Canadian public safety organizations apply a wide range of techniques and instrumentation (including the MMPI-2) to assess the probable behaviours, thoughts and emotions that public safety officer candidates may exhibit in the execution of their responsibilities.

Canadian Federal Public Safety. Federal public safety in Canada is provided by six major agencies: Canada Border Services Agency (CBSA), Royal Canadian Mounted Police (RCMP), Canadian Security Intelligence Service (CSIS), Canadian Security Establishment (CSE), Correctional Service Canada (CSC), and the Parole Board of Canada (PBC). The federal agencies responsible for public safety services are coordinated by Public Safety Canada (PS) and reviewed by Civilian Review and Complaints Commission for the RCMP (CRCC), the Office of Correctional Investigators (OCI) and the RCMP External Review Committee (ERC). This group of agencies is collectively responsible for the national security of Canadian citizens, border strategies, countering crime, and emergency management. Each department employs Canadians for a wide range of responsibilities relevant to the maintenance of peace and order in Canadian society ("About Public Safety Canada", 2016).

CSC. Federal correctional services administered by the CSC include the management, organization and treatment of incarcerated Canadians for crimes with sentence duration greater than two years, less a day. CSC front-line staff is categorized into three primary positions: correctional officer, primary worker and parole officer. Correctional officers maintain the safety and security of federal penitentiaries by monitoring, supervising and interacting with offenders. Correctional officers are tasked to watch for signs that the safety of others or security of the institution may be at risk. After establishing a level of risk for security concerns, correctional officers are required to act using appropriate security measures. Given that prison security is a 24/7 endeavour, correctional officers are required to work in shifts and certain officers are equipped with a firearm ("Correctional Officer", 2016). Primary workers perform similar roles as correctional officers but primary workers do not carry a firearm and work exclusively in federal women's facilities ("Primary Worker", 2016). CSC correctional officer and primary worker

recruits are evaluated for personality characteristics to ensure their suitability for these positions. Candidates are assessed via the completion of a MMPI-2 questionnaire, and, depending on resultant profiles, may be asked to participate in a structured interview ("Application Process", 2016).

Six Dimension Assessment Model

A division of the Public Service Commission of Canada (PSC), the Personnel Psychology Centre's staff of clinical psychologists were contracted by CSC to perform evaluations of correctional officer candidate psychological suitability from September 2012 to September 2015. The team assigned to conduct the assessments selected the MMPI-2 as the appropriate measure to identify behaviour patterns and psychological characteristics that are problematic in a corrections workplace. A consulting clinical psychologist, Dr. Denis Lapalme, provided a protocol and assessment model using the MMPI-2 in a content focused approach developed over a period of 15 years, from 1992 until 2007, and used to conduct the assessment of Ontario Provincial Police (OPP) candidates. The assessment procedure developed by Lapalme was applied by the PSC to determine the psychological suitability of approximately 1880 correctional officer/primary worker candidates for CSC.

Six dimension assessment procedure (2012-2015). Candidates were provided with standardized consent forms to acknowledge that their psychological data was to be used in their assessment process. In keeping with the standards of the College of Psychologists of Ontario, the Public Service Commission retains and archives candidate data for up to ten years following the assessment with both physical and digital files. Upon completing the MMPI-2 questionnaire, candidates were classified into one of three initial outcomes by the lead clinical psychologist.

These outcomes consisted of: no concerns, some concerns (requiring an interview), or severe concerns. If some concerns were raised by the candidate's profile, candidates would participate in a semi-structured clinical interview to investigate any concerns further. Interviews targeted one of six assessment dimensions: substance use, aggression, rigidity, cynicism, anxiety, and suicidal ideations. The six dimensions were developed in the assessment of provincial police officers. Interviews were assigned to a candidate in a priority order. First, the primary concern would be addressed, followed by additional concerns.

Candidates would be given a final decision (recommended or not recommended) based on a clinical interpretation of psychological data. The clinical interpretation was aimed at ruling out the presence of concerns, which might seriously affect the candidate's ability to perform the roles and responsibilities of a federal correctional officer. A formal report detailing the clinician's rationale, along with all interview notes, was stored with the candidate's test data. In the rare cases where the lead clinician was unable to render a conclusion, the PSC presented the case to a Suitability Review Committee for deliberation. The SRC was comprised of the Director of Security, Director General of Human Resources, Director of Recruitment, a representative from the legal department, and the lead clinician. Committee members presented additional information gathered during other phases of the hiring process to supplement findings from the psychological data. Additional information could take the form of previous employment history, security review, and substance use history. The SRC would continue to deliberate until a suitability decision was reached with unanimous agreement. The SRC decision and rationale was recorded in case the candidate made a request to know the justification for the recommendation, in accordance with the federal Act for Access to Information (ATIP). Should a candidate wish to

receive feedback, the lead clinical psychologist provided general information about the process, general decision criteria, and some specific concerns that had been flagged.

SIOP principles. The Society for Industrial and Organizational Psychology (SIOP) published a report in 2003 outlining a series of principles to assist in the validation and use of personnel selection procedures. The 2003 Principles Report aimed to collect and present established scientific findings and accepted professional practices to assist in the development, evaluation and use of psychological evaluations. This critique evaluates the six dimension model within SIOP's principles, highlighting strengths of, and possible improvements to, the content-focused approach for assessing psychological suitability of public safety officers. SIOP principles were applied as a measure of comparison to identify strengths and weaknesses of the content-focused approach.

Six dimension assessment model. The six dimension assessment model used to assess candidates in the correctional officer/primary worker selection process from September 2012 to September 2015 was based on the model developed by Dr. Denis Lapalme. This model of public safety psychological suitability was developed through OPP workplace observations and candidate MMPI-2 psychometric data. Through a series of Skype interviews, Dr. Lapalme provided a detailed account of the six dimension model development and implementation at the OPP.

Over 15 years, Lapalme validated associations between psychological data and officer outcomes at all stages of employment, recruitment, training, and within the community. The final product was a working model outlining the behavioural and psychological characteristics required for effective provincial police work. The assessment model was evaluated and validated quantitatively using a sample of officers' scales and critical item MMPI-2 response patterns in

contrast to a sample of Special Forces Officers. Special Forces Officers were selected as a validation comparison group for their high level of demonstrated job performance as police officers. Lapalme's six-dimension assessment offers unique contributions to the practice of psychological suitability assessments.

A content-focused approach is defined as one where assessment psychologists systematically review candidate psychological data for scale elevations as broad indicators, and, where there are significant elevations, explore the candidate's endorsed content responsible for the scale elevation. In parallel, assessment psychologists also review endorsed critical item lists that contain content that is problematic enough to be considered outside of the scale that it loads on. Test content is considered critical when 6.6% or less of the comparison group endorse an item in this manner. The definition of critical extends to items with contextually problematic content in relation to the assessment model content that may be more frequently endorsed. The nature of the content-focused approach deviates from traditional assessments of psychological suitability. Traditional assessments rely on scale elevations to inform decision-making on candidate psychological suitability (Fishler, 1997) and attempt to screen out candidates based on indicators of psychopathology. However, the content-focused approach is focused on assessing specific behaviours as indicators that influence performance on basic job requirements. The six dimension model is an effective example of a content-focused approach with a well-supported tool, outlining assessment dimensions that are core psychological characteristics for multiple public safety positions.

Assessment Model Strengths. The MMPI-2 was appropriate for this assessment model at the time of the model's conception. The MMPI-2 is, historically, a well-established instrument for psychological suitability assessments (Bartol, 1991; Kornfeld, 1995; Selbourn, Fischler and

Ben-Porath, 2007). The MMPI-2's 567 items allow for an effective delivery of the content-focused assessment model, covering a wide range of psychological phenomena. MMPI-2 content and supplementary scales contain infrequently endorsed questions. The forced choice True-False style of the questionnaire has a 4th grade reading level requirement and is valid across cultures (Graham, 1990). Canadian citizens come from many cultures, with 20.6% of Canadian citizens identifying a language other than French or English as their native tongue (Linguistics of Canadians, 2015). Competitions for national public safety positions are open to all Canadian citizens above the age of seventeen. It is therefore important that the tool used to identify suitable candidates for federal public safety positions is both accessible and unbiased to all Canadian citizens. A list of MMPI-2 scales and their description are included in Appendix G.

Assessment Model Weaknesses. Since the conception of this model there have been new and further validated tools developed. The original MMPI was first published by the University of Minnesota Press in 1943 (Shiele, Baker and Hathaway, 1943), the updated version of the MMPI was published in 1992 as the MMPI-2 (Hathaway et. al, 1989). Since 1989, representatives of the University of Minnesota have refined and restructured item and scale content to better reflect tool applications (one of those being public safety selection). The MMPI-2RF (Restructured Form) is better suited to evaluating public safety populations (Kitaeff, 2011). The MMPI-2RF scales are homogeneous and designed to more precisely measure distinct symptom constellations or disorders. These scales were developed with advanced psychometric theory, test development methods and statistical analysis that were not available when the MMPI-2 was developed (Ben-Porath and Tellegen, 2011). There are other additional measures that have established a level of predictive validity and internal reliability that provide content that is not covered in the MMPI-2. Multi-Health Systems, a Canadian organization, has developed a

police selection inventory, the Matrix-Predictive Uniform Law Enforcement Selection Evaluation (M-PULSE). The M-PULSE was designed to permit for law enforcement agencies to assess certain job performance risks. The M-PULSE uses two scales to assess profile validity; impression management and test attitude. The validity scales in this test are used to identify overly positive responding and candidates with a negative test taking attitude. The M-PULSE assesses candidates on 18 misconduct areas to measure behaviours associated with officer misconduct. The 18 misconduct areas range from general (Interpersonal Difficulties) to very specific (Inappropriate Use of Weapon). In addition to markers of validity and misconduct, candidates are assessed on attitudes, values and beliefs that have a direct relevance to police work. Empirical attitude scales assess candidates for negative self-attitudes, negative perceptions related to law enforcement, unethical behaviour, and unpredictability. The M-PULSE also supports the California POST (Peace Officer Standards and Training) assessment criteria with 10 POST scales. POST scales include measures of Social Incompetence, Recklessness, Rigidity, Poor Service Orientation, and Substance Use. The M-PULSE was validated to predict approximately 86% of liability cases (Davis and Rostow, 2017). Liability was defined as any of 16 possible events causing embarrassment or putting the organization at a disadvantage. Predictive validity for the M-PULSE was demonstrated by associating criterion variables with any of 16 liability concerns including: procedural/conduct mistakes, at fault motor vehicle accidents, weapons discharges, inappropriate weapons use, substance use, etc. (Williams, Davis, and Rostow, 2011).

Assessment Tool Weaknesses. The content focused approach deviates from current literature, which supports reliance on the use of clinical scales for selection decisions. Current literature supports reliance on clinical scales to render psychological suitability decisions due to

scales demonstrating robust reliability, lower face validity, and predictive validity (Sellborn et al., 2007; Ben-Porath and Tellegen, 2011). Predictive validity of the assessment model was not established and evidence supporting the relationship between scores and predicted events was not established. Predictive validity is the degree to which test scores predict performance on some future criterion (Shultz and Whitney, 2005). Predictive validity can be established by conducting a study of performance in the workplace of candidates who have continued working for the CSC and were evaluated by the PSC between 2012 and 2015. High correlations between the original measure and criterion variables integral to federal corrections work would reinforce the conclusion that the tool is a valid predictor of the specified criteria (DeVon et al., 2007). The six dimension model would see a benefit to implementing updated tools such as the MMPI-2RF (Kिताeff, 2011) or the M-PULSE (Davis and Rostow, 2017) in the stead of the MMPI-2. Further research would look to acquire evidence based on response processes in psychological suitability. It is well known that candidates in a selection process attempt to interpret the intent of the question, resulting in the restriction or alteration of their responses in candidates' efforts to present themselves as aligned with their image of the expected 'ideal candidate'. The effects of these response patterns can be better understood in the future by surveying test takers about their response strategies post-offer, analyzing examinee response times, and conducting experimental studies where the response set is manipulated.

Development. In a series of Skype interviews occurring in the month of September 2016 with Kevin Kennedy (author), Dr. Denis Lapalme answered questions relating to the development and validation of his six dimension content-focused approach. In the early stages of development, Lapalme recognized that there are significant differences between using the MMPI-2 tool for clinical assessment rather than for psychological suitability. Lapalme

established that the traditional uses of clinical and validity scales are not appropriate for the selection of suitable public safety officers. In his initial attempts to assess OPP officers, prior to his establishment of the model, Lapalme reviewed candidate profiles for the typical psychosomatic V pattern (depression clinical scale slightly elevated, with hypochondriasis and hysteria significantly elevated) with an additional elevation on Psychopathic Deviate (Pd) scale. From the feedback of recruitment staff and officers working in the field, Lapalme found that interpreting candidate profiles in this manner failed to identify problematic psychological characteristics for police work.

Lapalme thus adapted his assessment model to focus the assessment on content scales, supplementary scales, and items infrequently endorsed (critical items). Lapalme derived six primary dimensions from problematic scales and items. The dimensions were based on MMPI-2 content and were identified as substance abuse, aggression, rigidity, cynicism, anxiety and suicidal ideations. Traditional psychological suitability assessments during this era used the MMPI-2 clinical scales to come to a decision about the probable presence of psychopathology (Bartol, 1991). Lapalme surmised that in a selection context, MMPI-2 clinical scales do not measure psychopathology, and for this population, high scores on these scales should be interpreted as indicators of general distress.

Lapalme used the six dimensions as a basis for semi-structured follow-up interviews, where the assessing psychologist could identify a) the presence of a problem, b) the extent of a problem, and c) the coping mechanisms that inhibit performance as an officer. Items with a low response rate and critical content were regularly identified as true positives in an interview. For illustrative purposes, a sample of a critical item similar but not identical to those on the MMPI-2 is “I am afraid of who I’ll run into when I am shopping”. Concern would arise if the candidate

answered this item in the positive. Items such as this one have comparison group response rates less than 6.6%, so that a positive answer would be investigated in an interview. When candidates were faced with a specific example of a problematic response, they were asked to elaborate on their rationale for responding to that item. Candidates would readily acknowledge they had selected that response. The assessing psychologist would interpret the candidate's response and rationale when making the final decision of psychological suitability. Focusing the interview on the candidate's endorsed content held each candidate accountable for their responses. All dimensions apparent in candidate psychological profiles were evaluated by investigating the content supporting elevations on the content scale(s), supplementary scale(s) and critical item(s). The assessing psychologist accounted for all information present in the profile and determined which (if any and in what order) semi-structured interviews were appropriate for each profile. Throughout the development stage, OPP recruitment and operations staff offered insight into the importance of the dimensions and related problematic thoughts, behaviours and emotions. Workplace feedback permitted Lapalme to adjust the assessment model over time. The six assessment dimensions (anxiety, substance use, etc.) are used as the structure of the candidate interviews. The definition of each of the six assessment dimensions, how it is evaluated using the MMPI-2, and workplace feedback are discussed next (Lapalme, personal communication, September 2016).

Substance use. The first of the six dimensions is substance use. Candidates are evaluated by the scales: MacAndrews Alcoholism – Revised (MAC-R), Addictions Admissions (AAS), and Addiction Potential (APS) to determine current substance abuse habits or potential for developing substance use problems. Candidates scoring between 1.5 and 2 standard deviations above the mean on substance use related scales had APS, MAC-R and AAS item content

reviewed. Candidates scoring above 1.5 standard deviations and endorsing items suggesting substance use were selected for interview. The substance use semi-structured interview was designed to compare the candidate's alcohol use with Canada Health Drinking guidelines as a benchmark of excessive consumption, and, in addition, to review lifetime illicit substance use. Candidates demonstrating acceptable levels of substance use openly admitted to small amounts of drinking and presented a confident image of self-regulation. Candidates with difficulty regulating substance use behaviour tended to use substances for self-medication and stress management. Workplace feedback supported the conclusion that it is important for successful candidates to display a high level of self-regulation. Feedback also indicated that a regular practice of excessive alcohol consumption was not acceptable within OPP officer populations. The associated risks with hiring a candidate with a substance use problem as determined by OPP staff were: showing up late to work, showing up intoxicated to work, sick day abuse, and repeated absences for treatment to reduce substance use.

Aggression. Problematic aggression is evaluated from the candidate's Anger scale (ANG), Aggression scale (AGGR), and infrequently endorsed indicators of aggressive behaviours. Candidates scoring between 1.5 and 2 standard deviations above the mean on Aggression related scales had ANG and AGGR item content reviewed. Candidates demonstrating acceptable levels of aggression admit to feelings of anger and do not allow their feelings to directly affect their behaviour, therefore demonstrating effective emotional regulation. Candidates with difficulty regulating angry emotions express unregulated behavioural outbursts as normal reactions to challenging situations. Workplace feedback identified that officers expressing aggressive traits have an increased tendency to abuse power, which results in the reception of frequent complaints from the public. In high stress situations, officers with

aggressive behavioural patterns may become emotionally overwhelmed to the point that they neglect their own safety or the safety of others. Candidates with high scores on these scales are deemed to pose a significant risk to themselves and to colleagues in highly emotional, provoking situations, and present an unacceptable risk to the organisation.

Rigidity. High L and K scores were related to the selection context rather than to the candidate providing false information on the MMPI-2. Lapalme determined that there are other possible interpretations of what the L scale may be capturing and decided that it was not useful to conclude that elevated L and K scales were a consequence of the candidate lying. Instead, Lapalme introduced an interview to assess for high naivety or rigidity, a compromising trait for a police officer (Kappeler et al., 1992). Avoiding the assumption that candidates were disingenuous, Lapalme identified rigidity or guardedness as a personality style. Candidates presenting a personality style of rigidity or guardedness consistently had difficulties recognizing even the most common ‘inappropriate’ behaviours or faults in themselves. Candidates were selected for a rigidity/guardedness semi-structured interview based on positive responses to two of three questions that load on the L scale and are infrequently endorsed. A similar example to the items that would indicate a ‘rigid’ profile is a candidate responding ‘True’ to “There are many ways to do things, but only one right way”. Candidates demonstrating acceptable levels of rigidity can identify personal fault and are able to convey humane compassion to the public. Candidates with problematic rigid personality styles express an inability to admit common shortcomings and maintain standards for ethical principles as absolutes. Workplace feedback identified that officers who cannot admit fault are problematic; these officers are often inflexible or unwilling to compromise. Rigid officers have difficulty functioning in a team environment,

present social inflexibility, are characterized as highly judgemental and critical, and in extreme cases may be rejected or ostracized by fellow officers.

Cynicism. Cynicism is evaluated by the Cynicism scale (CYN), the Anti-Social Practices scale (APS), Negative Treatment scale (TRT) and cynicism related MMPI-2 critical items. Candidates scoring greater than 1.5 standard deviations above the mean on cynicism related scales had CYN, APS and TRT item content reviewed. Candidates with acceptable cynicism admit to negative perceptions of others but are readily able to identify with positive perceptions of others. Candidates with problematic cynicism maintain absolute negative perceptions and feel as though others are against them in life. Workplace feedback identified that officers with a cynical disposition have misgivings about others and have difficulties functioning in a team environment. Cynical officers present difficulties adopting the prosocial attitude of the organization. As officers, cynical candidates are likely to find scheduling ‘not fair’ and maintain a poor rapport with the public. Cynical officers are highly critical and contribute to an atmosphere of disrespect. In Lapalme’s group, Special Officer Recruits were actively excluded on cynical characteristics due to their demonstrated inability to cooperate.

General distress. Levels of general distress are assessed by Content Anxiety (ANX), Supplementary Anxiety (A), Low Self-Esteem (LSE), Social Discomfort (SOD), Fears (FRS), and critical items. Candidates scoring 1.5 standard deviations above the mean on anxiety related scales had ANX, A, LSE, SOD, FRS, and critical item content reviewed. Candidates demonstrating acceptable levels of anxiety admit to feelings of tension or worry but do not allow their feelings to direct their behaviour, demonstrating effective emotional regulation. Candidates with difficulty regulating anxious or distressful emotions become withdrawn and are unable to manage distressful or challenging situations. High anxiety officers demonstrated difficulty

engaging with others. Workplace feedback identified candidates who are self-critical, self-doubting and overly concerned with themselves as being at risk of becoming paralyzed in tense situations. Recruits with moderate to high anxiety levels had difficulty completing firearms training and maintaining their certification.

Suicidal ideation. Lapalme had yet to encounter a candidate with suicidal ideations; however, it is important to consider the possibility that a candidate can pose a risk to himself or herself. There are six questions on the MMPI-2 related to suicide. If any of the six questions are endorsed, the candidate is interviewed as soon as possible, and if necessary, is referred to appropriate resources.

Validation. As an additional form of validation, Lapalme used a comparative population of ‘Special Force’ officers to contrast MMPI-2 profile differences between hand-selected officers and the general population of officers. Officers recruited for these ‘Special Force’ teams were carefully selected by the department chief and senior officers to exceed expectations of regular duty performance. Comparing the group means and standard deviations, the Special Officer group’s content scales had smaller standard deviations and lower means. The Special Officers reported a lower endorsement rate on a large set of critical items highly relevant to the assessment process. Behaviours, thoughts and emotions identified in the critical item content by the assessing psychologist were, in turn, confirmed with officer recruiters as problematic. Continued workplace feedback supported that candidates with these concerns posed unacceptable risk to the institution, fellow officers, and the public (Lapalme, personal communication, September 2016).

Development strengths. The assessment parameters were developed and refined through a workplace feedback process. This was a long-term investment that resulted in refined

dimensions highly relevant to public safety work. Employees offered their experience of the work requirements of OPP officers and these were matched to the needs of the organization. The six dimensions were developed from the experiences of officers in the field. The six dimensions and its criteria were chosen based on work relevance, were clearly stated, and were acceptable in the social and legal context of the organization. Model criteria effectively represented individual outcomes such as work-related behaviours, outputs, attitudes, or performance in training. Officers providing information in the validation effort were knowledgeable and qualified for the criteria-related content they were asked to contribute. The criterion measures were developed in direct reference to the work performed. Assessment qualities were identified by careful consideration of the work setting and of stated organisational objectives. This allowed for reliable and relevant job information that is linked to anticipated behaviours and activities (Principles, 2003). The validation effort contrasting constables with high performance 'Special Force' officers inductively demonstrates the efficacy of this use of the MMPI-2 in identifying officer candidates who express psychological content that is problematic.

Development weaknesses. The six-dimension model did not demonstrate criterion validity, as suggested by SIOP principles (Principles, 2003). The criterion validity for assessing police recruits relied on the self-reported information provided by recruiting officers. The information provided by each officer contains a subjective account of valued qualities. The six dimensions were derived from the interpretation of recruiter expressions. The six dimension model did not establish a systematic quantitative validation of the six dimensions, or make a direct connection between assessment dimensions and the MMPI-2. In the future, this may be resolved by surveying subject matter experts and conducting a confirmatory factor analysis, anticipating six central factors. Subject matter experts should be identified as unbiased and

qualified subject matter experts (SMEs). SMEs should have thorough knowledge of work behaviours, activities, and responsibilities of the job incumbents and the knowledge, skills, abilities and other characteristics prerequisite to effective performance on the job. The survey for SMEs should meet the requirements of sound inference with: representative sampling, appropriate analysis techniques, and controls over plausible confounding factors. Attitudes of surveyed experts can also be accounted for to decrease criterion contamination from bias by standardizing collection procedures (Principles, 2003).

Conclusion

Public safety officers play an integral role in upholding Canadian laws and regulations. Personality assessments are important for the selection of public safety officers who are mentally capable of the roles and responsibilities of the work. It is important that these assessments are valid, unbiased and consistent, especially when recruiting candidates for federal positions.

Strengths of the six dimension model. The centralized approach of the six dimension content focused assessment allowed for control of evaluations, where final decisions were rendered by a clinical expert. In terms of SIOP's principles for assessment standards, the six dimension assessment model would benefit from a new tool and further validation efforts to support the six primary dimensions. At the time of model development, the MMPI-2 was the appropriate tool. The MMPI-2 provided a method of effectively delivering the content-focused approach and had demonstrated validity across cultures. In the model's construction, the workplace feedback process offered insight into highly applicable constructs with ties to real-world applications. The dimensions are relevant to public safety work, are clearly stated, and meet the needs of the organization. The dimensions were associated with individual outcomes and connected the MMPI-2 to anticipated behaviors and activities.

Weaknesses of the six dimension model. New tools such as the MMPI-2RF and the M-PULSE have stronger evidence of validity and reliability (Davis and Rostow, 2017; Kitaeff, 2011) than the MMPI-2 and were designed with psychological suitability assessments in mind. These tools provide more insight into officer liability by asking targeted questions that are addressed with the MMPI-2. The six dimension model development through a feedback process over time does not meet current standards for establishing construct or criterion validity. Current standards establish that assessment model and test criterion relationship should be established by a panel of experts or through quantitative methods (such as confirmatory factor analysis). The six central constructs were associated with the MMPI-2 by a single rater and no quantitative measures were employed. The model, being developed from the subjective reports of on duty officers, was not conducted within a systematic process of recording and interpreting information. The model used scales with high face validity, and did not use the established clinical scales. The six dimension model was not traditionally tested in terms of predictive validity. The predictive validity that was established was restricted to predicting selection for the higher performance ‘Special Forces’ group. This is problematic because it applies a ‘screen-in’ logic to a ‘screen-out’ process. Candidates identified as ‘high performers’ eligible for the special forces team were selected based on their merit, or, positive qualities. The fact that special forces officers had less critical items endorsed, lower means and restrictive standard deviations was not determined to be the driving factor for their selection into the high-performance group. The predictive validity that was established also did not account for officer liability outcomes or poor behaviour.

Overall, the six dimension model offers a depth of insight, not traditionally garnered, using scale elevations and predefined critical items. The approach is rational and holds

candidates accountable for their responses. However, the model should be administered through an updated tool and further efforts should be made to establish construct and criterion validity. Finally, the evidence establishing relationships between test scores and other variables of the six dimensions should be derived from quantitative methods.

Chapter 2 - Classification Algorithm: Discriminant Function Analysis

Introduction

The content-focused six dimension model of assessing psychological suitability contains an element of subjective appraisal for candidate evaluations. Because of this, additional variance in suitability recommendations can be attributed to psychologists' training and to their experience with the psychological suitability model. Psychologists' clinical experience has been shown to moderate clinical judgement and decision making when comparing graduate students and licensed psychologists (Waltman, Williams, and Christiansen, 2013). These differences can be applied to clinical psychologists who have no experience assessing candidates for psychological suitability. Daily clinical experience performing treatment and making clinical decisions is independent of rendering psychological suitability decisions. It is important that candidates are assessed consistently across Canada in federal selection processes for public safety positions (Employee Equity Act, 1995 c. 44). In the six dimension model, candidate psychological profiles are initially evaluated to determine whether the candidate meets the requirements, requires further review in an interview, or is unsuitable for public safety work. Candidates requiring further review are indicated for one or more of the six dimension-related interviews: substance use, aggression, rigidity, cynicism, anxiety, and suicidal ideations. Due to the content-focused nature of the model, it is difficult to categorize profiles by relevant concerns and clarify the importance of each concern. The difficulty of categorization is due to the magnitude of data and interpretative nuances of item content. Clinical decision support information systems link observations with expert knowledge to help decision makers make better choices (Garg et. al., 2005). Introducing clinical decision support information has the potential to improve the consistency and efficiency (Berner, 2007) of these evaluations among

assessors with various degrees of experience in the six dimension model. The discriminant function analysis is a statistical analysis used to predict a categorical dependent variable based on one or more continuous variables (Poulsen and French, 2008). In this chapter, a discriminant analysis was completed to determine if primary profile decisions could be determined from MMPI-2 profile data, which is to develop a method of estimating a profile's likely primary decision in the context of the six dimension model.

A discriminant analysis (Huberty, 1975) was conducted to determine a linear combination of variables with appropriate weighting to maximally discriminate between several primary outcomes of the six dimension model. This analysis aims to identify an accurate algorithm (weighted discriminant function) for predicting interview outcomes. All cases from the Public Service Commission dataset of correctional officer recruits were included in the analysis. A quadratic discriminant function analysis was conducted using an expert psychologist's decision as the dependent variable. All scales present in the dataset provided by the Public Service Commission were included in the analysis to allow for the highest degree of predictive accuracy of the linear combination.

Procedure

Population. The Public Service Commission provided selection data consisting of scales and item responses for 1413 candidates who had applied for a federal correctional officer position with Correctional Services Canada between September 2012 and September 2015. Selection data included a primary interview, hiring decision, MMPI-2 scale raw scores, and MMPI-2 item responses.

Data cleaning. Initially, 1413 correctional officer MMPI-2 profiles were provided by the PSC. All assessments that were cancelled at some point during the assessment process and did not have a final decision made on the profile were removed from the data set (n= 88). Assessments identified for a suicide interview (n=5) were all recognized as false positives during a follow-up interview and constituted an insufficient sample size to include as a ‘suicide interview’ outcome in this analysis. They were also removed from the data. A sample of 1320 profiles remained in the analysis.

Data. Forty-five MMPI-2 Scales provided in the PSC data set were used as predictor variables for this analysis. MMPI-2 scales included in the analysis are measures presented as standard in the MMPI-2 Interpretative Report, and they are grouped into five broad categories; validity, clinical, content, supplementary, and psy-5 (Appendix G). The discriminant analysis applied a “forced entry” method for variable entry; all variables were entered in a single step. The sample of candidate assessments contained seven possible outcomes. Candidates had been determined by expert assessors to pass, fail, or be further assessed in one of five interviews. Within the sample of candidate assessments provided, approximately 37% of cases had passed, 6% had failed and 57% had been selected for interview (5% substance use, 7% aggression, 6% rigidity, 7% cynicism, and 31% general distress). While during the real assessment candidate profiles were often identified for multiple (up to four) interviews, in this analysis of the data, the ‘primary concern’ or first interview indicated was used as the categorical dependent variable. For the analysis, group membership values were specified as proportional to the prior probabilities listed above.

Outcomes. MMPI-2 profiles measure a wide array of dimensions. Candidates were often considered for more than one interview to address all concerns in the profile. For the purposes of

this discriminant analysis, only the primary outcome was considered. The primary outcome is defined as the first interview indicated by the expert assessor in the dataset. Outcomes were assigned an alphabetical key; A: Pass, C: Fail, D: Substance Abuse Interview, E: Aggression Interview, F: Rigidity Interview, G: Cynicism Interview, H: General Distress/Anxiety Interview.

Statistical Analysis. Group centroids (vectors of means) on the two new canonical variables were calculated by applying the discriminant function weights to each case and computing the mean for each group of cases. The centroid for each group or level is defined as that group's average discriminant score on that function. Quadratic discriminant function analyses make assumptions about the distributions of the independent variables. Typical requirements are ratio level data, with multivariate normality and variance/covariance matrices of the variables that are homogenous across groups (McLachlan, 2004). Since the independent variables in the current context are psychological scale scores, they do not meet the requirements of ratio level data and multivariate normality criteria. MMPI-2 content, clinical and supplementary scales applied to a non-clinical population are positively skewed (Greene, 2011).

From a frequency analysis of MMPI-2 validity scales applied to this selection context, it was determined that they are negatively skewed. In an additional frequency analysis of MMPI-2 content, supplementary and psy-5 scales, it was determined that these scales were positively skewed. The validity scales were negatively skewed as is common with suitability selection processes, because candidates applying for a position tend to try to present themselves as highly suitable for the position. This contextual component leads to a negatively skewed distribution, opposed to Greene's 2011 statement for non-clinical populations. The heavily skewed data (both positive and negative) violates the assumption of multivariate normality. The goal of performing discriminant analysis using this psychological data is to discriminate between the seven possible

selection outcome decisions; the test of the efficacy applying this discriminant procedure in this context is not how well the data fit the assumptions of the technique, but how well the procedure solves the classification problem at hand. With this dataset, the large sample size (n= 1413) of candidate profiles assisted in mitigating skew problems of the discriminant function analysis. Heavily skewed data would violate the assumption of data normality. Without an exceptional sample size and multiple independent predictors, the classification algorithm would fail to accurately identify categorical outcomes (McLachlan, 2004).

Results

Canonical functions. The discriminant function analysis predicted seven outcome categories for psychological suitability decisions. To maximize discrimination between groups, six functions were generated. Eigenvalues describing variance explained by each function and likelihood ratio significance tests for zero eigenvalues were conducted, where the null hypothesis was that the canonical correlations are zero (Huberty, 1975). The collective group of functions were evaluated with four tests of significance. Tests of significance included Wilks' Lambda, Pillai's Trace, Hotelling-Lawley, and Roy's Max Root. All tests of significance rejected the null hypothesis, confirming that the group of canonical functions are significant when utilized together to predict group membership (Table 2.1).

Table 2.1

Collective Group of Functions test of Significance.

Test	Value	F	df	p
Wilks' Lambda	0.20	8.51	222	<.0001***
Pillai's Trace	1.24	7.18	222	<.0001***
Hotelling-Lawley	2.29	10.43	222	<.0001***
Roy's Max Root	1.54	42.41	37	<.0001***

Note. Overall tests for group of function's significances. The canonical r between function weight and groupings.

The top five contributing scales were selected from each of the canonical functions to describe how the functions apply the scores to discriminate among groups, although all scale weightings were used in the classification algorithm itself. The top five scales are highlighted in Table 2.2, which also describes all the functions and reports on group centroids and standardized canonical discriminant function weights.

Function 1. Function 1's five primary contributing scales are Anger (0.30), Defensiveness (0.26), Fears (0.25), Schizophrenia (0.23), and Anxiety (0.21). Group centroids suggest that when these scales are significantly elevated, a Fail outcome will be predicted ($M_{\text{fail}} = 4.11$). Low scores on Function 1 also serve to identify the profile as not belonging to a Pass outcome ($M_{\text{pass}} = -0.97$) or a Rigidity Interview outcome ($M_{\text{rigidity}} = -1.06$). Function 1 is extremely powerful, explaining 68.74% of the variance in the relationship between profile scores and outcomes ($p > 0.0001$).

Function 2. The five primary contributing scales of Function 2 are: Anger (-0.63), Addictions Admission Scale (-0.36), Fears (0.33), Trauma (-0.31), and Anxiety (Supplementary) (0.28). High scores on Function 2, which are associated with high scores on Fear and Anxiety and with low scores on the negatively weighted scales, predict that the profile will not belong to the categories of Substance Use Interview or Aggression Interview. Function 2 explains 11.54% of the variance of the relationship between profile scores and outcomes ($p > 0.0001$).

Function 3. The five primary contributing scales of Function 3 are: Anger (0.36), Over-controlled Hostility (0.27), Schizophrenia (0.26), Fears (-0.5), and Hypochondriasis (-0.26). When scores for Function 3 are elevated, they predict a Rigidity Interview outcome ($M_{\text{rigidity}} = 0.91$). Function 3 explains 8.00% of the variance of the relationship between profile scores and outcomes ($p > 0.0001$).

Function 4. Profiles with high scores on Function 4 would have some combination of high scores on Social Introversion (0.69), Defensiveness (0.57), and the Addictions Admission Scale (0.47), while having low scores on Anger (-0.60), Social Discomfort (-0.54). High scores on Function 4 predict a Substance Abuse Interview ($M_{\text{substance}} = 0.93$) and low scores predict an Aggression Interview ($M_{\text{aggression}} = -0.80$). Function 4 explains 5.05% of the variance in the relationship between profile scores and outcomes ($p > 0.0001$).

Function 5. Function 5 is constituted of high scores on the Low Self-Esteem (0.37) scale and the MacAndrews Alcoholism scale (0.35) and low scores on scales of Anti-social Practices (-0.85), Superlative Presentation (-0.40), and Psychological Distress (-0.32). A high score on Function 5 predicts a Substance Abuse Interview ($M_{\text{substance}} = 0.50$) while a low score predicts a Cynicism interview ($M_{\text{cynicism}} = -1.05$). Function 5 explains 4.32% of the variance of the relationship between profile scores and outcomes ($p > 0.0001$).

Function 6. The five primary contributing scales of Function 6 are: Frequency (0.49), Family Problems (-0.49), Depression (0.45), Marital Distress (0.40), and Anxiety (0.40). When these scales are elevated, Function 6 can effectively discriminate Rigidity ($M_{\text{rigidity}} = 0.66$) and Substance Use ($M_{\text{substance}} = 0.38$) Interview outcomes. Function 6 explains 2.35% of the variance of the relationship between profile scores and outcomes ($p > 0.0001$).

Table 2.2

Standardized Attributes of the 6 Canonical Functions that Predict Psychological Suitability Outcomes from Squared Scale Scores.

<u>Canonical Function Significance Tests</u>						
	Function 1	Function 2	Function 3	Function 4	Function 5	Function 6
Eigenvalue	1.57	0.26	0.18	0.12	0.10	0.05
% of Variance Explained	68.74	11.54	8.00	5.05	4.32	2.35

	Function 1	Function 2	Function 3	Function 4	Function 5	Function 6
p	<.0001	<.0001	<.0001	<.0001	<.0001	0.0002

Group Centroids

Pass	-0.97**	-0.01	0.17	0.03	0.14	-0.19
Fail	4.11**	0.11	0.74	0.22	0.14	-0.18
Substance Use Interview	0.16	-0.97**	-0.55	0.93**	0.50**	0.38*
Aggression Interview	0.44	-1.28**	0.16	-0.80**	-0.04	0.19
Rigidity Interview	-1.06**	0.75	0.91**	0.03	-0.07	0.66*
Cynicism Interview	-0.20	-0.39	0.02	0.43	-1.05**	-0.08
Anxiety Interview	0.48	0.40	-0.48	-0.15	-0.03	0.05

Standardized Canonical Discriminant Function Coefficients

Key	Scale	Validity Scales					
F	Frequency	-0.03	0.07	0.19	0.01	-0.19	0.49**
S	Superlative Presentation	0.08	0.1	0.19	-0.03	-0.4**	0.04
K	Defensiveness	0.26**	-0.12	0.16	0.57**	0.11	-0.15
Clinical Scales							
Hs	Hypochondriasis	0.02	0.06	-0.26**	0.26	-0.17	-0.12
D	Depression	0.01	0.1	0.09	0.02	-0.02	0.45**
Hy	Hysteria	0.01	0.12	0.13	-0.27	0.14	0.17
Pd	Psychopathic Deviant	-0.12	-0.02	0	-0.08	-0.06	-0.21
Pa	Paranoia	0.1	0.06	0.09	0.21	-0.1	0.05
Pt	Psychasthenia	-0.07	0.02	-0.22	-0.09	0.06	0.36
Sc	Schizophrenia	0.23**	0.07	0.26**	0.19	0.24	0.02
Ma	Hypomania	0.04	-0.01	0.06	0.07	-0.08	0.12
Si	Social Introversion	0.05	0.12	0.25	0.69**	0.18	0.09
Content Scales							
ANX	Anxiety	0.21**	0.26	0.05	0.01	0.05	-0.13
FRS	Fears	0.25**	0.33**	-0.5**	-0.19	0.23	0.17
OBS	Obsessions	0.04	0.15	0.03	-0.2	-0.07	0.21
DEP	Depression	0.13	0.11	0.17	0.03	-0.03	-0.03
HEA	Health Concerns	0.08	0.07	-0.14	-0.14	0.04	0.08
BIZ	Bizarre Mentation	0.15	0.24	0.14	0.03	0.05	0
ANG	Anger	0.3**	-0.63**	0.36**	-0.6**	0.01	0.21
CYN	Cynicism	0.14	0.02	0.03	0.41	-0.23	-0.17
ASP	Anti-social Practices	0.12	0.14	-0.1	0.02	-0.85**	-0.13
TPA	Type-A Personality	-0.02	0.16	-0.1	-0.06	-0.06	0.06

		Function 1	Function 2	Function 3	Function 4	Function 5	Function 6
LSE	Low Self-Esteem	0.15	-0.25	0.17	-0.02	0.37**	-0.18
SOD	Social Discomfort	0.02	0.04	-0.19	-0.54**	-0.18	-0.07
FAM	Family Problems	0.07	0.02	0.12	-0.06	0.14	-0.49**
WRK	Work Interference	0.2	-0.04	0.23	0.04	0.09	-0.1
TRT	Negative Treatment	0.01	-0.05	0.13	0.2	-0.16	-0.1
Supplementary							
A	Anxiety	0.15	0.28**	0.03	0.16	-0.07	-0.4**
Mt	Maltreatment	0.09	-0.05	0.09	-0.06	-0.07	0.2
Pk	Trauma	-0.13	-0.31**	-0.11	-0.02	0.06	-0.36
MDS	Marital Distress	0.02	-0.11	-0.21	0.13	0.14	0.4**
O-H	Over-controlled Hostility	0	0.05	0.27**	0.1	0.03	0.12
MAC-R	MacAndrews Alcoholism	-0.12	-0.14	-0.13	0.38	0.35**	0.23
AAS	Addictions Admission Scale	0.15	-0.36**	-0.22	0.47**	0.3	0.29
APS	Addictions Potential Scale	0.02	-0.05	-0.2	0.44	-0.19	-0.34
PSY-5							
AGGR	Aggression	0.09	-0.14	0.12	-0.1	0.02	0.27
PSYC	Psychological Distress	-0.09	-0.23	-0.05	-0.3	-0.32**	0.2

*Note. **Group centroids**, the mean discriminant scores were calculated for each outcome for all six functions. The farther apart the means are, the less error there is in classification. **Standardized scoring coefficients** are partial correlations that are representative of each predictor's unique contribution to each function. These coefficients are indicative of the relative importance of each predictor in predicting the outcomes from each function. Important predictors discussed in text are marked in bold type face with **.*

Classification. In this assessment model, the human decision maker has an inherent level of error when interpreting complex psychological data. Since the discriminant analysis is based on this imperfect data, this affects discriminant analysis classification, the generated group centroids, and the matrix structure. The inconsistency from human error directly contributes to the misclassification rate, where the classification algorithm attributes the greatest probability to a different outcome than the expert. When the classification algorithm was applied to the PSC data, the algorithm misclassified 175 of 1320 cases included in the analysis (13.25%, Entropy $R^2 = 0.67$). A classification table of outcomes is available in Table 2.3. The algorithm correctly

matched the outcome decision of the expert for Fail outcomes most frequently, correctly predicting 96.34% of cases. The Substance Use Interview outcome was correctly selected in 88.41% of cases, the Aggression Interview for 84.38% of cases, the Rigidity Interview for 94.94% of cases, and the Cynicism Interview for 83.52% of cases. The algorithm performed poorest in identifying Anxiety Interview outcomes, correctly identifying 79.95% of cases. Cases classified as any Interview outcome by the expert assessor were occasionally incorrectly classified as Pass outcomes by the discriminant analysis (Substance Use, 8.70%; Aggression, 11.46%; Rigidity, 5.06%; Cynicism, 13.19%; Anxiety, 16.09%). Cases classified as Pass outcome by the expert assessor were most often incorrectly classified as Anxiety Interview (6.6%). The interviews most commonly mistakenly classified as a Pass were Anxiety and Cynicism interviews.

Table 2.3

Case Classification Results for Classification Algorithm on PSC Psychological Suitability Cases.

		Expert Outcome (Suggested Interview)						
		Pass	Fail	Substance	Aggression	Rigidity	Cynicism	Anxiety
Predicted Outcome	Total Cases	499	82	69	96	79	91	404
	Pass	90.18%	0.00%	8.70%	11.46%	5.06%	13.19%	16.09%
	Fail	0.40%	96.34%	0.00%	0.00%	0.00%	0.00%	1.24%
	Substance Use	0.20%	0.00%	88.41%	0.00%	0.00%	1.10%	0.74%
	Aggression	0.60%	1.22%	0.00%	84.38%	0.00%	0.00%	0.50%
	Rigidity	1.80%	0.00%	0.00%	0.00%	94.94%	0.00%	0.74%
	Cynicism	0.20%	0.00%	0.00%	1.04%	0.00%	83.52%	0.74%
	Anxiety	6.61%	2.44%	2.90%	3.13%	0.00%	2.20%	79.95%

*Note. Overall classification accuracy for the algorithm is 86.75%, percentages reflect proportions of the outcome columns that were accurately identified. Entropy $R^2 = 0.67$, $p < 0.0001$ ****

Overall, the results of the multivariate analysis of the case sample provided by the PSC show that, the psychological scales included in this quadratic discriminant analysis result in an 86.75% accuracy of correctly predicting the expert decision.

Conclusion

A discriminant analysis was conducted to determine a linear combination of variables with appropriate weighting to maximally discriminate between several primary outcomes of the six dimension model to predict profile classification from profile data. The discriminant function analysis determined the initial psychological suitability decision from six canonical functions. Function 1 accounted for most the variability in the relationship between scale scores and outcome categories. Function 1 operated to identify cases for Pass, Fail and Rigidity Interview. Functions two to six added sensitivity to the classification algorithm to identify Substance Use Interview, Aggression Interview, Cynicism Interview, and Anxiety Interview outcomes. For these other interview outcome categories, the accuracy rate when the classification algorithm was reapplied to the dataset was less accurate, but suitable for the intended application. It was determined that the quadratic discriminant classification function effectively classified most cases.

For the six dimension assessment model, the variance in predictive accuracy may be attributed to the content focused approach. In addition to evaluating scale elevations, the assessing psychologist interprets endorsed content and decides upon an outcome from infrequently endorsed, problematic content that may not be reflected by an elevated scale. This classification algorithm is applied specifically for identifying the likely outcomes of psychological suitability within the content focused six dimension model and can operate adequately to assist clinical decision making processes for public safety officer candidates for this model. The classification algorithm's overall performance indicates a consistent method in interpreting these types of psychological suitability decisions, despite the subjective nature of the content focused approach. The classification algorithm can be implemented to moderate clinician

experience (Berner, 2007) such that federal hiring decisions improve in consistency. Human judges and the computer algorithm both rely on MMPI-2 scales to reach decisions. The computer algorithm's ability to make decisions was similar to how a human judge would decide on most cases.

Chapter 3: Clinical Decision Support Systems for Psychological Assessments

Introduction

Clinical Decision Support Systems (CDSS) offer an additional resource to decision makers who are interpreting clinical information (Berner, 2007). Psychologists employed by the Public Service Commission use complex psychological data to determine candidate suitability for public safety positions. CDSS implemented into selection processes can further support Public Service clinicians while interpreting complex psychological data. Federal selection processes often employ multiple regional assessors to administer psychological evaluations for public safety positions. This is done for the assessment process to be accessible to the candidate. In Canada, requirements for becoming a registered psychologist or psychotherapist vary among provinces. In a public safety assessment process conducted by the Public Service Commission of Canada for Correctional Services Canada, regional psychologists were contracted to administer assessments and conduct necessary follow up interviews. Over the duration of the assessment process for correctional staff (2012-2015), candidate profiles were interpreted by a lead clinical psychologist. The lead clinical psychologist position that requires advanced training in psychometric theory and statistics, and its primary responsibility involves making the initial decision on the suitability of public safety candidates (Principles, 2003). The role of the lead clinical psychologist is an active one that involves reviewing satellite psychologists' reports, interfacing with the public safety department, providing feedback to candidates, and determining

an initial decision on the applicant's profile and a final decision after an interview. When the lead clinician is unavailable, or has changed roles within the department, it can dramatically affect the consistency of the evaluations. Staff clinical psychologists temporarily act as the lead clinical psychologist when the lead clinician is on vacation, ill, or on extended leave. Interim clinical leads have varied experience conducting these types of evaluations and render decisions by their own decision processes. To provide equal opportunity among public safety candidates (Employee Equity Act, 1995), it is important that clinical decisions be rendered in consistent fashion, despite changes to the person making these decisions. Clinical decision support systems can be implemented to support the temporary clinical lead in navigating the psychological data such that candidates are evaluated more consistently between assessors. A clinical decision support system implementation might also act to reduce the human error (Keen, 1987) of the clinical lead, improving the overall integrity and consistency of the assessment (Berner, 2007).

Theoretical Background

Clinical decision support systems. Decision support systems (DSS) were initially developed in 1976 to improve effectiveness of decision makers without intruding on autonomy. Decision support systems were introduced as management solutions for semi-structured business tasks (Keen and Morton, 1978). With advances in computing technology integrating multiple data sources (Keen, 1987), DSS applications were implemented for use outside semi-structured business tasks. Clinical decision systems (CDSS) are information systems designed to improve clinical decision making. In medicine, current patient data is applied to previous patient data with sophisticated algorithms that generate patient-specific recommendations. Recommendations generated from these algorithms can suggest treatment options for infertility (Ehrig et. al, 2016), diagnose chest pain (Kong et. al, 2012), or assist in identifying breast cancer (Aruna,

Nandakishore and Rajagopalan, 2012), along with many other applications. From a systematic review of the literature, it has been noted that many CDSSs improve practitioner performance, although effects on patient outcomes are understudied and inconsistent (Garg et. al, 2005). In clinical psychology, clinical performance has been compared to mechanical prediction; where given data from a predictive model is contrasted with a practicing clinician's ability to identify the correct outcome. In a meta-analysis, Grove and colleagues (2000) identified that mechanical prediction is typically as accurate or more accurate than clinical prediction. Psychological diagnoses are defined by a means of referral to normality, where a sub population behaves differently from the general populace. Over time, the criteria for diagnoses are adjusted within a culture, as new diagnoses are introduced or old ones are modified. Mechanical prediction relies heavily on rigid definitions and post-data relationships to generate outcomes. Although an exclusively mechanistic approach for psychological disorders leads to recursive clinical decisions and does not capture a changing culture, mechanical algorithms can still provide support in clinical psychology practice, especially if they are updated regularly. Clinicians with access to estimated outcomes generated from mechanical algorithms can use the additional information to support their conclusions while also exercising autonomy in decision-making processes.

CDSS for psychological suitability. Public safety candidates are evaluated on their personality traits, behaviour patterns and psychological characteristics that may be problematic in the workplace. The qualities of psychological suitability assessments identify these assessments as an ideal framework to study the implications of CDSS in psychology.

Psychological suitability assessments have a limited number of possible outcomes, are performed by clinical psychologists and often use measures of psychopathology. Large assessment

processes collect data from thousands of applicants where the response sets and decisions can be analyzed to support future decisions. The PSC provided data from a correctional officer psychological suitability assessment process that spanned from 2012 to 2015. The outcomes associated with each MMPI-2 profile were assessed within a content-focused six dimension model developed by Dr. Denis Lapalme, where candidates were assessed on the need for interview focused on: substance use, aggression, rigidity, cynicism, anxiety, and suicidal ideations. Dr. Lapalme developed his model of public safety psychological suitability through OPP workplace observations and candidate MMPI-2 psychometric data. The assessment model was evaluated and validated quantitatively using a sample of officers' scale results and critical item MMPI-2 response patterns in contrast with a sample of Special Forces officers. In the PSC dataset, correctional officer candidates were determined to pass, require interview(s), or fail. Candidates with an interview outcome were identified with a primary interview, along with any additional interviews. In the second chapter of this thesis, PSC data were analyzed using a discriminant function analysis to develop the classification algorithm that predicted outcomes based on MMPI-2 data. The classification algorithm was reapplied to the data, and it generated probability estimates of the 7 possible outcomes. The probability estimates are the decision support system clinicians participating in this study used to assist in their decision processes. Half of the profiles selected for experimentation contained the CDSS information with an estimated likelihood of the possible profile decisions generated from the algorithm described in the second chapter.

Selecting profiles for CDSS evaluation. MMPI-2 psychological profiles contain 567 True-False statements that are the basis of any psychological suitability interpretations from a psychologist. Candidates from a normal population have the potential to select a wide range of

item endorsement combinations that is reflective of an individual's thoughts, behaviours and emotions. Psychologists participating in this experiment were asked to assess a total of fourteen candidate profiles, with two groups of seven where the outcome estimates were either presented or not presented. To compare match accuracy between groups, the fourteen profiles were required to be similar enough such that psychologists would not require additional time or demonstrate reduced accuracy because of extreme profiles. If the predicted outcomes also ranged in clarity, this could differentially affect psychologists' ability to decide on the profile. Clarity, in this context, refers to having multiple probability estimates of likely outcomes that do not discriminate effectively between outcomes. This experiment is concerned with how psychologists apply decision support information when provided with a clear outcome. Two measures, abnormality and entropy, were introduced to sample profiles for use in the experiment such that the profiles selected would not introduce variability of profile and decision difficulty across interview types.

Abnormality. To control for the level of interpretation required from psychologists for varied profiles, a weighted score was generated to reflect a profile's total number of infrequent deviant responses. The score was weighted by the infrequency of the deviant response ($1-p$).

Shannon's entropy. Information theory studies the quantification, storage, and communication of information (Shannon, 1948). Information theory was developed to understand how information was communicated from a transmitter through a channel to a receiver. A key measure of uncertainty in the transfer of information is Shannon's entropy (H). Shannon's entropy serves as a measure of the unpredictability of the state, or its average information content.

$$H = - \sum_{i=1}^k p_i \log_2 p_i$$

H is the amount of information that is gained when one learns the outcome of a random process (Shannon, 1948). H is calculated by the negative sum of the probability of each possible outcome times the log of the probability of each possible outcome (above). H is considered the amount of information that is gained once the actual outcome is known. Maximal uncertainty about what will be selected occurs when all probabilities are equally likely. In this experiment, Shannon's entropy was applied as a measure of uncertainty of each profile's set of probabilities for the seven possible outcomes. A high H indicates that the classification algorithm has generated less conclusive results; the highest possible H would have all the possible outcomes with the same estimated likelihood. H represents how much information the assessor receives from the CDSS, or how successfully the CDSS help assessors narrow down their choices.

Methods

Securities, data storage and data export. An online application was developed to collect data from psychologists in a task where they employed scales based on the MMPI-2 to conduct 14 psychological suitability evaluations from profiles. The online application provided flexibility to participants, allowing participants to choose to complete the study at their convenience as the target population of psychologists were not available for scheduled participation at one specific time or place. Participants were provided with randomly generated unique access keys for a secure instance of the application. The link was provided to participants via email along with general instructions to contact the study administrator if they experienced any technical difficulties. The data was stored in private online web server hosted in Toronto Ontario, with an SSL certificate. Participant data was stored on the server and was accessed

through an export link. The export link requires the user to know the link address, the administrator username, and the administrator password as authentication to retrieve participant data. Each of the participating psychologists was directed to their participant link, and watched a nine minute brief video outlining the task and the assessment model. After watching the brief video, participants were asked to consent to their participation and filled out a short survey outlining their experiences assessing candidates for psychological suitability. Upon the completion of the questionnaire, psychologists were presented with one profile at a time, and asked to make and confirm their decision in response to each case. Finally, after completing the experiment, psychologists were provided a debrief outlining the intent of the experiment and were directed to additional resources.

Instructions, consent, and experience survey. A nine minute brief video outlining the task of making psychological suitability assessments for correctional officers using a five-dimension approach (Appendix A). The original six dimensions were reduced to five because the PSC dataset had no cases identified for the sixth dimension (suicidal ideations). This prevented the discriminant analysis from effectively generating scoring coefficients for the interview type. The brief video featured a series of visual aids (slides) accompanied by audio commentary. A CONTINUE button allowed for participants to proceed to the next step once the brief content was reviewed. Participants had the ability to return to the brief video at any time during the experiment. A timing feature was incorporated that that paused time recording during mouse and keyboard inactivity (five seconds) and navigation away from the profile (e.g. leaving the tab or browser). Participants were informed of the intent of the study and the use of their data (Appendix B). Ethics approval was obtained by Laurentian's Research Ethics Board (Appendix A). Participants were provided with the information related to use of their data, and were told

what to do if any ethical concerns arose. Participants were warned that they might encounter a mild level of stress while making profile decisions. Participants were asked four short questions to approximately quantify their experience conducting psychological suitability assessments before proceeding to the experiment (Appendix C).

Participant population. Registered, currently practicing, clinical psychologists were recruited to participate. Contact information was collected from provincial registries and psychologists from all provinces were contacted via telephone for participation. A group of Public Service Commission staff and contract psychologists from the CSC 2012-2015 assessment process (n=8) participated as experienced assessors. A group (n=5) of registered clinical psychologists participated as inexperienced assessors.

Procedure. Participants were presented with fourteen correctional officer applicant MMPI-2 profiles, one at a time, and asked to use profile response sets to render a decision as to the applicant's psychological suitability for work in a penitentiary as a correctional officer. Half (seven) of the fourteen profiles contained an estimate of the individual profile's likely outcome based on the discriminant function analyses described in the second chapter of this thesis. The profiles were presented in a random order for each participant, to control for order effects. After reviewing the critical items, scales, and the content, participants were asked to choose one of seven outcomes for each profile. Participants could decide between: Pass- profile contained few or irrelevant concerns, Interview- profile contained concerns that should be addressed in one of five interviews, and Fail- profile contained many highly problematic concerns (see Table 3.1).

Table 3.1

Psychological Suitability Decision Outcomes

Outcome	Description
Pass – A	Candidate has indicated few or little problematic content, suitable for correctional work
Fail – C	Candidate has indicated excessive or damning problematic content, unsuitable for correctional work
Substance Use Interview – D	Candidate profile indicates recent and/or current use of illicit substances, excessive use of alcohol, or prescription medication abuse that should be investigated in an interview.
Aggression Interview – E	Candidate profile indicates hostility, rage, anger, or will engage in physical confrontation. This candidate may break things or intimidate others and should be reviewed in an interview.
Rigidity Interview – F	Candidate profile indicates an inability to consider the viewpoint of others, and/or has difficulty admitting to social transgressions that should be investigated in an interview.
Cynicism Interview – G	Candidate profile indicates that they find people are untrustworthy, untruthful, uncaring and exploitive of others. Attitudes and beliefs should be further investigated in an interview.
Anxiety Interview – H	Candidate profile is indicative of a general or specific overactive fear response. Candidate presents as nervous and excessively worried about possible misfortunes. The candidate is anxious, easily frightened and should be investigated further in an interview.

Note. Outcome descriptions provide an overview of the assessment model criteria decisions. Further details of assessment criteria can be found in Chapter 1, on pages 19-24

Profiles. MMPI-2 profiles of correctional officer applicants assessed between 2012 and 2015 with a moderate level of abnormality ($1.57 > \log(x) < 2.22$, $M = 1.89$) and entropy ($-4.29 > \log(x) < -0.62$, $M = -2.30$) were selected to be included in this experiment (PSC). Shannon's Entropy (H) was used in this experiment to identify the level of information the clinical decision support system provided to participants. These profiles contained a variety of psychological characteristics to help the participant choose a best outcome out of the seven possible choices. Each participant was required to consider characteristics of the profile relevant to the six dimension model (excluding suicide), and then determine an outcome based on the characteristics important in correctional work. Profiles presented z-scores that measured the

extent to which each profile's scales were different from a normative comparison group. The comparison group used for scoring was a sample of 800 OPP officers (PSC DATA).

Endorsement frequency statistics were generated from the dataset of 1800 correctional officer applicants between 2012 and 2015 (PSC DATA). Scale scores were presented in the form of z-scores, providing the distance between the profile's scale score and the comparison group mean in standard deviations. Two profiles of each outcome were randomly sampled from the correctional officer applicants that met the criteria of moderate ambiguity and moderate abnormality.

$$Abnormality = \sum(1 - P)$$

$$Entropy (Shannon's H) = -\sum p * \log_2(p)$$

A “moderate” profile is defined as one falling within the 2nd or 3rd quartile of abnormality and entropy. Profile abnormality was determined based on the proportion of questions on the profile that were responded to in a manner deviant from the comparison sample, where the response (TRUE or FALSE) proportion was less than 30 percent. These items accounted for 469 of the 567 MMPI-2 items. Outcome probabilities were ranked and evaluated to determine if the top likelihood estimate from the discriminant function matched the expert decision. Profiles with a predicted outcome matching the expert decision were included in the sample. To correct for a negative skew of the distribution of entropy and abnormality scores, a log transformation was applied to the data. Controlling for abnormality is a crucial step when studying decision making processes on psychological profiles. Profiles with high abnormality offer an additional challenge to the participant, having more content to interpret and integrate into decision processes. Profiles with high entropy value present more conflicting information with decision support. A decision support system with multiple outcomes that were highly rated

would not lead the participant to a clear decision. Future studies may aim to investigate how entropy in a decision support system affects decision making processes.

Table 3.1

Profile Sampling Conditions to Control for Abnormality and the 'Information (H)' Conveyed by Probability Estimates

	<u>Condition for the Inclusion of a Profile in Sampling</u>
1. Log(H)	$-4.29 > x < -0.62$, $M = -2.30$
2. Log(Abnormality)	$1.57 > x < 2.22$, $M = 1.89$
3. Expert Decision	Match with predicted outcome from discriminate function analysis

Note. Shannon's H is a measure of 'information', profiles with multiple viable estimates may add an element of variability to the experiment and are therefore controlled for within the profile sampling procedure.

To test if the conditions for inclusion were biased for certain outcomes, sample proportions were evaluated to identify if they were significantly different from the complete data set using z-tests. The null hypothesis for each comparison was that the refined sample of profiles that meet inclusion criteria is not significantly different from the sample of profiles in the PSC dataset. Rejecting the null hypothesis would imply that the profile sampling conditions have biased the sample of experimental profiles. It was anticipated that profiles from the Pass profiles would be over inclusive and Fail outcomes would be restrictive due to the criterion restricting the range of profile abnormality as well as the range of outcome likelihood entropy. Restricting the range of profiles included for sampling to profiles within a normal range of abnormality systematically will restrict the sample Fail outcomes. Fail outcome profiles typically contain infrequently endorsed, problematic content and profile outcomes are associated with a low level of entropy; for Fail cases, the classification algorithm made consistent, confident predictions. Profiles with Pass outcomes contained frequently endorsed content. This resulted in a decreased abnormality score for the Pass outcome profiles; thereby resulting in a greater proportion of profiles that met sampling criteria when contrasted with the proportion of Pass cases in the PSC dataset. The purpose of this comparison was to determine a sampling bias for interview

outcomes. Bonferroni comparison for 7 comparisons set the $\alpha_{adj} = 0.00365$ for 2-tailed test, $z_{crit} = \pm 2.683$ at $\alpha = 0.05$ (Table 3.2).

$$p_{pool} = (p_1 * n_1 + p_2 * n_2) / (n_1 + n_2) \quad SE = \sqrt{p_{pool} * (1 - p_{pool}) * (\frac{1}{n_1} + \frac{1}{n_2})} \quad z = (p_1 - p_2) / SE$$

Table 3.2

<i>Z Hypothesis Test Determine if Cases that Meet the Sample Pool are Representative of the PSC Dataset</i>							
	<u>Pass</u>	<u>Fail</u>	<u>Substance</u>	<u>Aggression</u>	<u>Rigidity</u>	<u>Cynicism</u>	<u>Anxiety</u>
Count: Total	419	79	66	84	78	81	308
Proportion (%)	37.58	7.09	5.92	7.53	7.00	7.26	27.62
Count: Met Requirements	353	5	39	54	46	63	223
Proportion (%)	45.08	0.64	4.98	6.9	5.87	8.05	28.48
-	-	-	-	-	-	-	-
p	0.407	0.440	0.055	0.073	0.065	0.076	0.28
SE	0.023	0.01	0.011	0.012	0.012	0.012	0.021
z	3.277***	-6.723***	-0.88	-0.526	-0.973	0.633	0.409

Note. This method was applied to determine if sampling criteria favoured the inclusion of profiles in an outcome.

There were significantly more profiles that met the inclusion criteria in the Pass outcome ($z = 3.277$, $p < 0.001$) and fewer profiles that met the criteria for the Fail outcome ($z = 6.723$, $p < 0.001$). Sample differences for Pass and Fail outcomes are attributed to the restrictions placed on profile abnormality and entropy in the profile sample conditions (Table 3.1). Substance Interview ($z = -0.88$), Aggression Interview ($z = -0.526$), Rigidity Interview ($z = 0.973$), Cynicism Interview ($z = 0.633$), and Anxiety Interview ($z = 0.409$) accepted the null of containing no significant sample bias and therefore appropriate for random sampling to identify profiles for inclusion in the experiment.

Two profiles for each outcome were randomly selected from each outcome group of sample profiles. One of the two profiles for each outcome was randomly assigned a SHOWN or NOT SHOWN condition. A profile assigned to the SHOWN group was presented to participants

with the decision support. A profile assigned to the NOT SHOWN group was presented to participants without decision support. Participants were informed of the proportions of profile outcomes upon study completion. The efforts to control for abnormality and uncertainty were taken to avoid using extremely abnormal profiles in the experiment; this was done such that profiles could be randomly sampled from each of the outcome categories where the possible profiles had a clear outcome and a moderated amount of problematic content to interpret.

Profile interface. The content focused assessment model was considered in the development of the user interface. The experiment profiles were interactive to explore the depths of profile content. Each scale on the profiles could be selected to view the items contributing to the scale for that profile. When a scale was selected, participants were provided with item endorsement frequency beside item content and the profile's response of 'TRUE' or 'FALSE'. Scales with z-scores greater than 1.5 and less than 1.75 were coloured yellow as moderately elevated. Scales with z-scores greater than 1.75 were coloured red as significantly elevated. Item endorsement statistics were gradationally colour coded from white to yellow for endorsement percentages less than fifty and greater than five percent. Item endorsement statistics were gradationally colour coded from yellow to red for endorsement percentages from five percent to zero percent. This colour scheme was adopted to draw attention to significant scale elevations and infrequently endorsed content. To decide the outcome for the profile, participants would indicate their selection by ticking a box beside the appropriate outcome. Participants were limited to one response per profile. To continue to the next profile, participants had to confirm that their selection was complete (Appendix C: Profile Presentation).

Data. With the unique participant keys, if a participant accidentally or purposefully closed the browser they would be able to return to their last profile and time recording would

resume. Each decision and time to decision was stored with the profile and the user key. Time recording occurred from the instant the first profile in the experimentation component was presented to the participant in their browser. Time recording was paused if the participant navigated away from the page, or remained inactive for 5 seconds. After each decision, time to decision was recorded along with the decision of each profile. Difference times were calculated for each outcome, where the profile with outcomes SHOWN was subtracted from the profile with outcomes NOT SHOWN. Positive values indicated that the profile with outcomes SHOWN took less time than those where the outcomes were NOT SHOWN.

Profiles were associated with an expected outcome previously determined by an expert assessor during the original assessment process. Participant decisions were recorded. Profile decisions were contrasted with the expert decision, and assigned a 1 (match) or 0 (no match) value. All participants were scored out of 7 for SHOWN profiles and 7 for NOT SHOWN profiles. 14 points was the maximum for a measure of overall match accuracy. Participants were asked 4 questions related to their experience conducting public safety psychological suitability assessments. Participants were questioned on how many years they participated in an assessment process, how many profiles they had evaluated, the approximate proportion of interviews conducted for these profiles, and the approximate proportion of the profiles where they were responsible for the decision. The number of years participants conducted psychological suitability assessments was used as the covariate in the MANCOVA analysis.

Debrief. Upon completion of the study, the final page of the online application was a debriefing statement with additional resources of information related to the study. Participants were told exactly what the manipulation of the experiment was and where additional information on CDSS literature could be found. Participants were thanked and provided with information on

the REB ethics committee whom they could contact for any ethical concerns arising from their participation (Appendix D).

Results

A multivariate analysis of covariance (MANCOVA), an extension of analysis of covariance, allows for the statistical control of continuous independent variables that might significantly affect the dependent variables. MANCOVA was used to control varied assessor experience, which has been demonstrated to affect performance (Garg et. al, 2005). A 2x2 MANCOVA was performed with accuracy and time as the dependent variables and whether probability estimates for the outcomes were SHOWN or NOT SHOWN as a within subjects variable. Experience of the psychologist in years was entered as the covariate. Cohen's Partial Eta Squared ($p\eta^2$) was used as a measure of effect size. Richardson (2011) and Cohen (1969) identify $p\eta^2$ of 0.0099 as a small effect, 0.0588 as a medium effect and 0.1379 as a large effect. Experience was significant as a covariate for time ($F_{1,182} = 12.136, p = 0.001, p\eta^2 = 0.068$), but was not for accuracy ($F_{1,182} = 3.631, p = 0.058, p\eta^2 = 0.021$). For the main effect of outcome conditions SHOWN and NOT SHOWN, there were significant differences for accuracy ($F_{1, 182} = 8.230, p < 0.05, p\eta^2 = 0.047$). The main effect of time was not significant ($p = 0.46$). Using standard error bars on a mean plot, participants had a greater accuracy when they were presented with outcome likelihood estimates (Figure 3.1). When participants were presented with decision support, they improved their match accuracy from 49% to 68%. For the outcome variable, there were also significant differences for accuracy ($F_{6, 182} = 3.65, p < 0.05, p\eta^2 = 0.12$). Differences in performance are shown with two standard errors on mean plots of accuracy by outcome type (Figure 3.2).

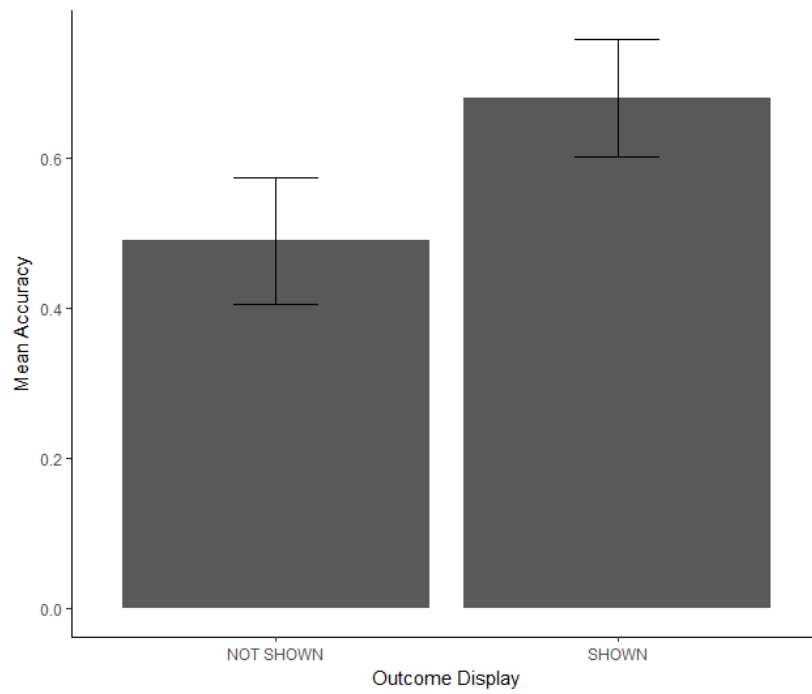


Figure 3.1 Overall accuracy mean plot for cases where outcome probabilities were NOT SHOWN/SHOWN.

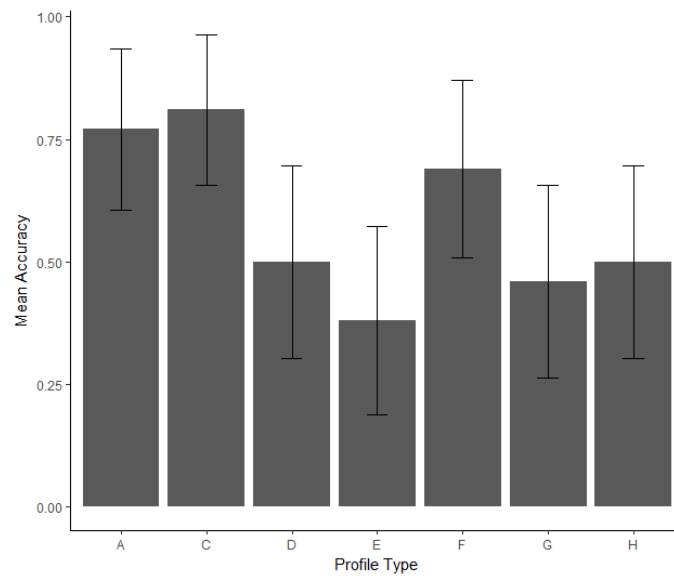


Figure 3.2 Mean accuracy plot for each profile type, A: Pass, C: Fail, D: Substance Use Interview, E: Aggression Interview, F: Rigidity Interview, G: Cynicism Interview, H: Anxiety Interview.

Paired t-tests were conducted to identify differences in match accuracy between profile types, testing at $\alpha = 0.05$. Participating psychologists matched the expert decision more often with the Pass than the Substance Use Interview ($t = 2.27, p = 0.032$), the Aggression Interview ($t = 2.32, p = 0.009$); and the Cynicism Interview ($t = 2.54, p = 0.018$). Participants matched the expert decision more often with the Fail than Substance Use Interview ($t = 2.309, p = 0.029$), the Aggression Interview ($t = 3.7336, p = 0.001$), the Cynicism Interview ($t = 2.807, p = 0.009$), and Anxiety Interview ($t = 2.54, p = 0.018$). There was a significant interaction between the type of profile and the outcome (SHOWN, NOT SHOWN) for accuracy ($F_{6, 182} = 4.575, p < 0.05, \eta^2 = 0.141$). Differences in accuracy are shown with two standard errors on mean plots of accuracy by outcome type for SHOWN and NOT SHOWN conditions (Figure 3.3). There was a significant difference between SHOWN and NOT SHOWN conditions for Aggression Interview, with participants matching the expert more frequently when provided with SHOWN outcomes ($t = 6.32, p = 0.0001$).

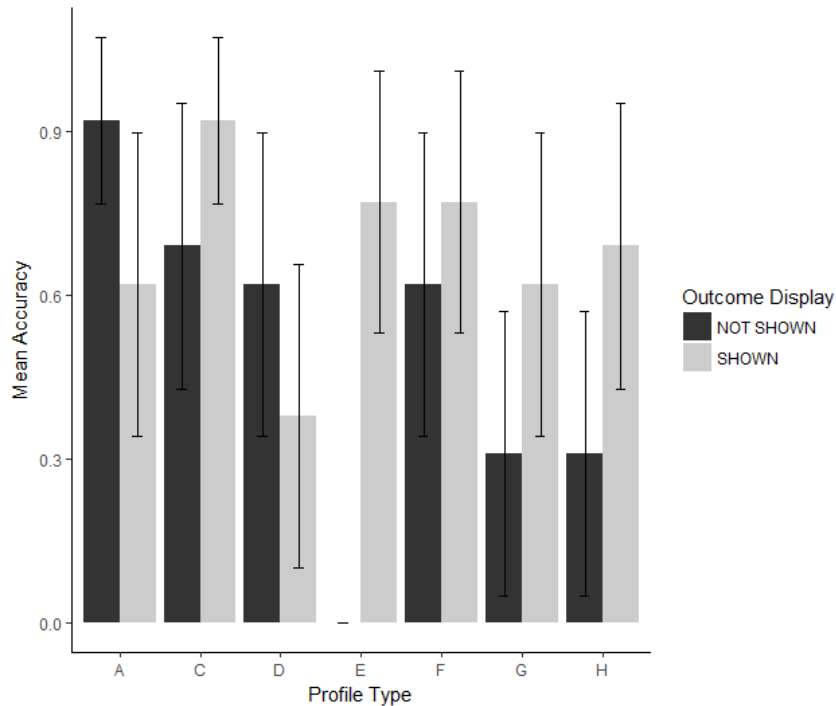


Figure 3.3 Mean accuracy for each profile type where the profile had an outcome Display: NOT SHOWN and SHOWN.

Limitations

This experiment was designed to investigate the ‘average’ profile; average is defined as within the inter-quartile range of abnormality and entropy. This is not an accurate depiction of a real-world candidate population, where an assessment process includes a wider range of uncertainty and abnormality in the profiles. Profiles are not typically presented in an online platform with colour mapping of key statistical information. The information arrangement may have a significant effect on how an assessor interprets the psychological data and statistical information. The ‘expert decision’ of the profiles included in the experiment were made at a time that afforded additional interviews to be indicated if the assessing psychologist saw fit. Due to the random sampling procedure, it is plausible that the profiles included in the experiment could be identified as requiring multiple interviews. This would affect the participating psychologist’s interpretation of the profile, leading to a decision that differs from an expert, but may still be

valid as a present concern in the profile. This assessment model did not establish a measure of predictive validity, associating assessment dimensions with organizational outcomes. Without an established level of predictive validity the decision support system is limited in its efficacy of determining suitable correctional officers within this model, using the MMPI-2. Model improvements with the consideration of protective/positive factors; e.g. resilience (Conner and Davidson, 2003), positive emotions (Fredrickson, 2004); as well as adjusting for the vulnerability-stress model. These factors can be integrated into this 6 dimension content focused model to further support the validity of the hiring decisions and thereby increasing the utility of the decision support in identifying the most suitable candidates. The decision support system provides an effective conduit to make quantitative adjustments based on protective factors to reduce the complexity of profile interpretation.

An additional limitation was that the classification algorithm was established to predict a single primary outcome without considering secondary outcomes when interviews are required. In the future, this limitation can be rectified in future designs with more data, where additional discriminant analyses can be established to predict any secondary or additional interviews. Within the PSC dataset of correctional officer applicants from 2012-2015, it was common for candidates to have multiple interviews. As the classification algorithm currently operates, a candidate's outcome estimates can indicate that the candidate may require more than one interview as the entropy of the outcome estimates increases. Implementing this type of decision support would have to be tested for its utility in improving clinician match accuracy on profiles with multiple interviews and would make for promising future research. In additional future experimentation, an analysis of the experimental profiles from a panel of expert assessors to determine inter-rater reliability for profile outcomes would be valuable. The six dimension

model is a non-traditional model of psychological suitability assessment that relies heavily on the interpretation of the content that generates the psychological data. A decision support system may prove much more effective in a model that relies less on interpretive judgement, and more on the scale elevations. The participant population ($n = 14$) consisted of registered psychologists currently practicing with basic training in measures of psychopathology located in British Columbia, Quebec, and Ontario. This is a very small sample of practicing registered psychologists in Canada, and therefore the results of the study cannot be generalized to the broader population of psychologists.

Conclusion

The online application with decision support was developed and implemented by Kevin Kennedy with assistance from Donald Harvey. This structured organization of MMPI-2 data provided psychologists with accessible information for profile scale elevations, profile content, item endorsement statistics, and estimates of expected outcomes to support psychological suitability assessment decisions. The online application is highly interactive and allows psychologists to identify scale elevations from highlighted values with visual cues for moderately and significantly elevated scales (relative to the comparison group). Psychologists are also able to explore candidate's endorsed content that resulted in scale elevations, allowing for rich information of candidate's psychological presentation. Feedback was received on the experiment from three participants expressing their thoughts on the decision support application for interpreting MMPI-2 data. No participating psychologists expressed negative views of the experiment user interface or decision support system.

“Having no experience in psychological assessment for staffing purposes (public safety positions or not), I was hesitant to participate, so it's a good thing to mention (or to take into consideration) that no

prior experience is mandatory. Once I decided to go ahead with it, I found it easy to understand, very well done and useful in the long term. Not sure I want to know if I got any of the profiles right though :).”

“I completed the experiment. I found it quite interesting and the instructions and format were really well done (ref. with a click we could access scales related items).”

“[The decision support system] looks great. It simplifies the decision making a lot. Especially with the probability window. I found the outcomes very guiding in reviewing the psychological data.”

This study provides insight into how regression based decision support systems for psychological decisions might be useful to psychologists who are acting as decision makers in a hiring situation for the PSC. When presented with SHOWN outcomes, this sample of assessors did not match the expert decision for every case ($A_{\text{SHOWN}} = 68\%$). This suggests that some psychologists did not blindly rely on the recommendation of the DSS and continue to exercise a level of independent judgement despite being presented with a viable outcome. Overall, decision support systems can be implemented to support a temporary clinical lead in navigating the psychological data such that candidates are evaluated more consistently for psychological suitability evaluations. Implementing a decision support system in practice can act as a guide for interpreting psychological data, reduce the error of both experienced and inexperienced assessors, and improve the integrity of the assessment.

Assessment processes that require a level of consistency among multiple assessors will benefit from implementing a classification algorithm into the assessment profile. Implementing the display of an index of outcome likelihoods will help less experienced psychologists assess candidates along similar lines as expert psychologists while they learn assessment model qualities. Additional outcome information may act as a guide to interpreting the profile, while still providing the assessor with the final decision. An index of outcome likelihood, which offers

the probabilities associated with each outcome as these are predicted by a discriminant function analysis provides assessors with more solid grounds to explore the profile for concerns, and can improve match accuracy from 49% to 68%. Providing an index of outcome likelihood does not affect the amount of time it takes psychologists to render a decision. Time was not identified as a significant effect in this study, implying that the psychologists used outcome estimates to guide their decision, and did not rely solely on the classification algorithm to determine profile outcomes. Pass and Fail outcomes matched expert decisions more frequently than Substance Use Interview, Aggression Interview, and Cynicism Interview, implicating that deciding between interview outcomes results in varied responses from psychologists.

Discussion

This analysis supports the conclusion that a decision support system does offer a level of support to the consistency of psychological suitability evaluations and is worthwhile to implement in practice. Improving the consistency between evaluations is a valuable trait when one considers the stake of hiring for employment where Canadian citizens expect to be evaluated fairly for federal selection processes. An outcome likelihood DSS does not have significant implications for candidate profiles with a clear outcome of Pass, Fail or 'Rigidity'. However, in determining appropriate interviews, presenting the psychologist with likelihood estimates does improve consistency regardless of experience levels. In this experiment, psychologists were briefed on the five dimensions used to evaluate the candidate profile. The Aggression Interview outcome profile that did not display outcome estimates performed very poorly, where almost no assessors matched the expert decision. In these cases, in which the decision for the profile is unclear, the assessment process will benefit in consistency from having a DSS implemented. When participants were faced with Substance Use Interview profiles, participating psychologists

were less inclined to agree with outcome recommendations and often went against the recommendations of the probability estimates that were displayed. It is surmised by this author that substance use is understood relative to personal experience, especially when substance related questions on the MMPI-2 are broad indicators of substance use. This fact may have affected how participants interpret levels of substance use, where any history of drinking to excess may be unacceptable from the eyes of one psychologist resulting in a Fail outcome, while admitting to historical drug use for another psychologist may be less indicative of a problem and seen as a natural phase of development, resulting in a Pass outcome. In both situations, substance use is evaluated by the relative severity of the items endorsed (of the candidate profile) and having access to the likelihood that it is a Substance Use Interview did not demonstrate a positive effect of assisting the participating psychologist in making a decision that matched with the expert.

In this experiment, it appears participants followed the recommendations of the DSS when they had also concluded their own decision, and that it was the appropriate recommendation for the profile. The results of this demonstrate that regression applications can be implemented into practice to influence the decision-making processes of psychologists conducting psychological suitability evaluations, and interactive applications can serve to support psychological decision making processes. In future studies, similar applications should be contrasted with traditional assessment practices to determine the effect of interactive applications in decision making for psychological assessments.

Appendices

Appendix A: Brief Script

Thank you for participating in this research study. During the next few minutes you will be guided through an assessment model and software application you will be using to render psychological suitability decisions. This experiment investigates decision making processes when selecting correctional officer candidates for psychological suitability. As a participating assessor, you will be presented with 14 candidate profiles. It is your role as the assessing psychologist to determine whether a candidate: is suitable for correctional work, requires further review in an interview, or appears unsuitable for correctional work. You make judgement on item content, item endorsement statistics, scale z-scores, and estimated outcomes to identify the most appropriate outcome for each case. There are only 14 profiles, and each should be considered thoroughly to make the best possible decision for the institution. Next, you will be briefed on assessment protocols and important information to consider with each decision. For this task, you will consider each MMPI-2 profile for the role of a federal Correctional Officer. Correctional Officers (COs) maintain the safety and security of federal penitentiaries. They monitor, supervise and interact with offenders. They regularly watch for signs that the safety of others or security of the institution might be at risk. When necessary, they take appropriate security measures. COs will conduct routine patrols and inmate counts, supervise inmate movement and escort inmates both inside and outside the institution. COs will also search cells, offenders, visitors, vehicles, living units, and surrounding areas. If necessary, COs will conduct security checks and perform other duties.

Other duties COs may perform:

- verify safety equipment
- draft daily logs
- submit reports
- brief visitors, volunteers, and other criminal justice professionals who enter the institution

Most law enforcement settings employ rigorous initial screening processes before candidates are subject to complete any measure of personality. This results in a refined selection pool where most pathological candidates are eliminated by earlier assessments. Most candidates at this stage tend to not have a criminal history or antisocial indicators. Candidate scales are measured against a comparison group of public safety officers. **Candidate z-scores demonstrate how different the candidate is from the comparison group.** Comparison group scales have smaller means and restrictive standard deviations, magnifying relatively smaller differences in profile presentations. If a scale is significantly elevated, it identifies statistical differences between the **candidate and the comparison group** and **does not** categorize the candidate as having clinically significant levels of that scale.

Initially, you be shown an overview of the candidate profile presenting all validity, clinical, content, and supplementary scales with the corresponding z-scores. Z-scores greater than 1.5 and less than 1.75 are indicated yellow as moderately elevated, z-scores greater than 1.75 are indicated red as significantly elevated and likely problematic. Negative z-scores indicate that the profile scores below the comparison group mean. Since this is a normal (non-clinical) population, it is likely to have profiles with negative z-scores and no items loading on a scale. Some profiles contain an **estimated likelihood for each outcome**. This estimate was generated from a discriminant function analysis and **accurately identifies the decision of an expert assessor 87% of the time**. You will have the option to review the scale content. These scales and items should be considered relative to 5 broad dimensions.

1: Substance Abuse

- Recent and/or current use of illicit substances
- Excessive use and abuse of alcohol
- Prescription medication abuse
- Active drug problem

2: Aggression

- Easily angered or upset, hotheaded
- Expressing anger directly in fights, breaking things and intimidating others

3: Rigidity/Guardedness

- Inability to take into consideration the viewpoint of others

- Difficulty admitting to social transgressions and faults most people willingly acknowledge.
- Rigidity interviews are determined from an item list with 3 indicators, if all 3 items are indicated, it is highly likely that the candidate requires an interview for rigidity.

4: Cynicism

- People are untrustworthy, untruthful, uncaring and exploitive of others
- Belief that others are only interested in themselves and will take advantage of them given the opportunity

5: General Distress

- General or specific reporting of an overactive fear response
- Nervous, excessive worry about possible misfortunes
- Anxious, easily frightened

Your decision as to the candidate's outcome should be supported within these 5 dimensions. It is up to you to **determine the most appropriate outcome for each profile**. You may only select **one outcome**, so choose wisely! You will be asked to choose one of the following several outcomes to select the most appropriate outcome for each case. Interviews are effective at determining the presence and extent of any concerns present in the profile. If you believe the candidate to have concerns that need further review before a final decision, select an interview outcome. Passing a candidate is stating that there is not enough problematic content to warrant an interview, and they are already suitable for correctional work. Failing a candidate is making a statement that the candidate has excessive problematic content and should not be considered further for a correctional officer position.

It is important to establish a familiarity with the application as these are complicated profiles with many possible outcomes. It is up to you as an assessor to determine the best possible outcome given the population of interest and the role of a correctional officer. This is a candidate MMPI-2 profile. The scales are grouped by their type; validity, clinical, content, supplementary and item lists. Each scale has a z-score, where the candidate's score is compared to a population of 1800 correctional officer applicants. Z-scores equal to or greater than 1.5 are indicated yellow. Z-scores equal to or greater than 1.75 are indicated red. By selecting the scale name, you can choose to review the scale content. As this model relies on the interpretation of candidate response sets, please take the time to review each candidate's responses for problematic content. Some profiles contain a display of that specific profile's likely outcome. The probability is applied from combining the data of 1800 correctional applicants, where expert assessors have made decisions for similar profiles. This algorithm is correct 87% of the time when classifying cases from within this population. Upon selection, you have access to the questions selected that load on the scale, this applicant's response to the question and the rate of the population for answering the question in this manner. If an applicant has indicated "True" and the corresponding response rate is 30%, 30% of correctional officer applicants surveyed endorsed this item as "True". Response rates are indicated from yellow to red depending on how uncommon the item's response is. Red items are very infrequently endorsed in that manner and likely require further review. It is important to note that there may be some overlap, where items load on multiple scales or indicators. Items in the critical item list may also be found in additional scales. However, some items on the critical lists are not found on the scales. Therefore, it is important to review critical items when they are indicated. Please take your time and exercise sound judgement on the implications of each response set to pass candidates with few or irrelevant concerns, interview candidates with concerns that deserve further review, and fail candidates with highly problematic concerns.

Appendix B: Consent Form

I, the participant, am interested in participating in this psychological study, conducted by Kevin Kennedy and supervised by Dr. Cynthia Whissell. The purpose of the study is to broaden the understanding of how additional statistical measures can be introduced to assist in clinician decision making.

If I agree to participate, my participation will consist of participating in a single 45-minute – 1 hour testing session conducted online. Initially, I will be briefed on a 5-dimension content focused approach to assessing public safety officer psychological suitability. After, I will assess 14 correctional officer candidate MMPI- 2 profiles for their psychological suitability applying the 5-dimension approach. For each profile reviewed, I will be presented with scale and item content information specific to each profile. As this is sensitive test data and information, I will not reproduce or copy any information provided in the experiment.

- My professional ability will not be evaluated or commented on in any form.
- No personal information will be disclosed, and complete anonymity will be maintained.

The information I provide by participating in this study may contribute my own understanding of personnel psychological suitability screening for public safety officers. This research is aimed to improve the efficiency and consistency of these types of decisions, and may support or detract from the use of classification algorithms to provide statistical information to support clinical decision making.

I have also received assurance from the researchers that the information I will share will remain strictly confidential. Electronic data will be kept for 10 years on an AES encrypted USB within a secured office.

If I have any questions or concerns, I may contact Kevin Kennedy or Dr. Cynthia Whissell. If I wish to address any questions or concerns with an official not attached to the research team regarding possible ethical issues or complaints about the research itself I may contact a Research Ethics Officer at the Office of Research Services at 705-675-1151 x 3681/2436 or toll-free at 1-800-461-4030, or by email ethics@laurentian.ca.

I understand that rendering suitability decisions may be stressful and challenging. If I am experiencing unusual stress, I will end my participation. My participation is strictly voluntary and I am free to withdraw from the study at any moment or refuse to participate without penalty.

Thank you for your participation!

☐

I accept and agree to participate

Appendix C: Experience Survey

Approximately how many years have you conducted public safety psychological suitability assessments?

(If none enter 0)

Approximately how many psychological profiles have you reviewed for public safety psychological suitability assessments?

(If none enter 0)

Approximately what percentage of those profiles were for the purpose of administering an interview?

(If none enter 0)

For approximately what percentage of those profiles were you responsible for the decision?

(If none enter 0)

Profile 3 of 14

Validity z-scores		Content z-scores	
Name	2	Name	2
Integrity	0.83	Anxiety	2
Lie	-1.20	Fears	0.63
Deftness	-1.33	Obsessions	1.88
Superstition	-0.30	Depression	1.13
		Health Concerns	-0.78
		Beane Thoughts	-0.07
		Anger	-0.17
		Cynicism	0.66
		Anti-social Practices	1.17
		Type-A Personality	0.79
		Low Self-Esteem	-0.89
		Social Discomfort	-1.19
		Family Problems	-0.30
		Work Interference	0.14
		Negative Treatment	0.08

Supplementary z-scores	
Name	2
Aggression	0.31
Anxiety	1.57
Alcoholism	0.84
Addiction Admissions	2.17
Addiction Potential	0.47

Item Tests	
Name	Count
Critical	8
Rigidity	0

Question		
Question	Response	% Endorsed
I have never been in trouble with the law	False	18.81%
I can remember "paying sick" to get out of school	True	53.72%
I can express my true feelings only when I drink	True	9.37%
People tell me I have a problem with alcohol, but I disagree	True	9.37%

Clinical z-scores	
Name	2
Hypochondriasis	0.74
Depression	-0.04
Hysteria	-0.08
Psychopathic Deviant	1.07
Paranoia	1.29
Psychasthenia	2
Schizophrenia	0.80
Hypomania	2.38
Social Introversion	-0.14

Outcomes	
Name	Probability
Anxiety	0.872
Fail	0.128
Pass	0.000
Rigidity	0.000
Substance Use	0.000
Cynicism	0.000
Aggression	0.000

Decision (profile 3 of 14)

Pass

Interview: Substance use

Interview: Aggression

Interview: Rigidity

Interview: Cynicism

Interview: Anxiety

Fail

Submit decision

Question		
Question	Response	% Endorsed
I have never been in trouble with the law	False	18.81%
I can remember "paying sick" to get out of school	True	53.72%
I can express my true feelings only when I drink	True	9.37%
People tell me I have a problem with alcohol, but I disagree	True	9.37%

Appendix E: Debrief

This study is concerned with how providing regression information with psychological suitability profiles affects clinical decision making. Medicine (and other fields) have implemented regression information into their decision-making processes. When clinicians have had access to this additional information, they perform less ‘errors’ and take more time to consider the information. This occurs especially when there are differences in levels of clinical experience with an assessment model. Clinical decision support systems are relatively new advancements to the medical field and have yet to be seen studied for psychological assessments.

How Was This Tested?

In this study, you were asked to decide an outcome for 14 candidate profiles. Of the 14 profiles, there were 2 profiles of each type of outcome. 2 passes, 2 fails and 2 of every type of interview. 7 of the 14 profiles contained an estimated likely outcome (described in the brief). Time and match accuracy were recorded and your experience conducting these types of evaluations was accounted for. Your actions in the module were recorded and reviewed for errors in time reporting and programming errors.

Hypotheses and Main Questions

We expected to find that when provided with the additional information, clinicians would take more time to render their decision. We estimated that the increased cognitive load would add time to the time between decisions on profiles with likelihood estimates. This effect is suggested to be moderated by the level of experience the participating clinician has in making these types of decisions. More experienced clinicians would benefit less from the regression information and clinicians with significant experience would not have improved match accuracy or an increased time of decision.

Why is This Important?

Federal government hiring processes require the assessment of large groups of candidates. These assessments often require the recruitment of clinicians with varied experience and training with assessment models and tools. This research may support the implementation of similar regression information to improve the consistency and efficiency of national assessment processes. Regression information may be useful in supporting decisions, instead of making the decision independent of any human input. This application of providing likelihood estimates uses the mechanisms of statistical learning to support clinical decision making in practice. If successful, this source of additional information can be applied to other areas of complex, clinical decision making.

What If I Want to Know More?

If you are interested in learning more about decision making processes and clinical decision support systems (CDSS), you may wish to consult:

Publications

- Kawamoto, K., Houlihan, C. A., Balas, E. A., & Lobach, D. F. (2005). Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Bmj*, 330(7494), 765.
- Berner, E. S. (2007). *Clinical decision support systems*. New York: Springer Science+ Business Media, LLC.
- Musen, M. A., Middleton, B., & Greenes, R. A. (2014). Clinical decision-support systems. In *Biomedical informatics* (pp. 643-674). Springer London.

If you have any concerns about your rights as a participant in this experiment, please contact a Research Ethics Officer at the Office of Research Services at 705-675-1151 x 3681/2436 or toll-free at 1-800-461-4030, or by email ethics@laurentian.ca.

Thank you again for your participation!

Appendix F: Ethics Approval



APPROVAL FOR CONDUCTING RESEARCH INVOLVING HUMAN SUBJECTS

Research Ethics Board – Laurentian University

This letter confirms that the research project identified below has successfully passed the ethics review by the Laurentian University Research Ethics Board (REB). Your ethics approval date, other milestone dates, and any special conditions for your project are indicated below.

TYPE OF APPROVAL / New <input checked="" type="checkbox"/> / Modifications to project / Time extension	
Name of Principal Investigator and school/department	Kevin Kennedy, supervisors , M. Lariviere, Human Kinetics, C. Whissell, Psych, N. Gingras
Title of Project	Supporting Psychological Suitability Decisions with Mechanical Prediction
REB file number	6009720
Date of original approval of project	Feb 7 th , 2017
Date of approval of project modifications or extension (if applicable)	
Final/Interim report due on: (You may request an extension)	February, 2018
Conditions placed on project	

During the course of your research, no deviations from, or changes to, the protocol, recruitment or consent forms may be initiated without prior written approval from the REB. If you wish to modify your research project, please refer to the Research Ethics website to complete the appropriate REB form.

All projects must submit a report to REB at least once per year. If involvement with human participants continues for longer than one year (e.g. you have not completed the objectives of the study and have not yet terminated contact with the participants, except for feedback of final results to participants), you must request an extension using the appropriate LU REB form. In all cases, please ensure that your research complies with Tri-Council Policy Statement (TCPS). Also please quote your REB file number on all future correspondence with the REB office.

Congratulations and best wishes in conducting your research.

Rosanna Langer, PHD, Chair, *Laurentian University Research Ethics Board*

Appendix G: Relevant MMPI-2 Scales

<i>MMPI-2 Scales and Descriptions</i>		
Key	Scale	Validity Scales
F	Frequency	Client "faking bad" (in first half of test)
S	Superlative Presentation	Improving upon K scale, "appearing excessively good"
K	Defensiveness	Denial/evasiveness
Clinical Scales		
Hs	Hypochondriasis	Concern with bodily symptoms
D	Depression	Depressive symptoms
Hy	Hysteria	Awareness of problems and vulnerabilities
Pd	Psychopathic Deviant	Conflict, struggle, anger, respect for society's rules
Pa	Paranoia	Level of trust, suspiciousness, sensitivity
Pt	Psychasthenia	Worry, anxiety, tension, doubts, obsessiveness
Sc	Schizophrenia	Odd thinking and social alienation
Ma	Hypomania	Level of excitability
Si	Social Introversion	People orientation
Content Scales		
ANX	Anxiety	General symptoms of anxiety, somatic problems, nervousness or worry
FRS	Fears	Specific fears and general fearfulness
OBS	Obsessions	Difficulty making decisions, excessive rumination and dislike of change
DEP	Depression	Feelings of low mood, lack of energy, suicidal ideation and other depressive features
HEA	Health Concerns	Concerns about illness and physical symptoms
BIZ	Bizarre Mentation	The presence of psychotic thought processes
ANG	Anger	Feelings and expression of anger
CYN	Cynicism	Distrust and suspiciousness of other people and their motives
ASP	Anti-social Practices	Expression of nonconforming attitudes and possible issues with authority
TPA	Type-A Personality	Irritability, impatience and competitiveness
LSE	Low Self-Esteem	Negative attitudes about self, own ability and submissiveness
SOD	Social Discomfort	Preferring to be alone and discomfort when meeting new people
FAM	Family Problems	Resentment, anger and perceived lack of support from family members
WRK	Work Interference	Attitudes that contribute to poor work performance
TRT	Negative Treatment	Feelings of pessimism and unwillingness to reveal personal information to others
Supplementary		
A	Anxiety	Generalized emotional distress, related to anxiety
Mt	Maladjustment	Generalized emotional distress relating to college maladjustment
Pk	Trauma	Post-traumatic stress disorder diagnostic scale - Keane
MDS	Marital Distress	Generalized emotional distress relating to marital dysfunction

O-H	Over-controlled Hostility	Behavioural dyscontrol, repression of hostile thoughts and emotions
MAC-R	MacAndrews Alcoholism	Correlates of substance users
AAS	Addictions Admission Scale	Behavioural dyscontrol relating to specific substance use/abuse behaviours
APS	Addictions Potential Scale	Preference for social engagements, outgoing, expressive correlates of substance users, behavioural dyscontrol
		PSY-5
AGGR	Aggression	Measures the accuracy of an individual's inner representation of objective reality, often associated with perceptual aberration and magical ideation.
PSYC	Psychological Distress	Measures an individual's tendency towards overt and instrumental aggression that typically includes a sense of grandiosity and a desire for power.

Note. Scale Descriptions (Greene, 2011; Pearson, 2015)

References

- Aamodt, M. G. (2004). Special issue on using MMPI-2 scale configurations in law enforcement selection: Introduction and meta-analysis. *Applied HRM Research*, 9(2), 41-52.
- Anderson, G. S., Litzenberger, R., and Plecas, D. (2002). Physical evidence of police officer stress. *Policing: an international journal of police strategies and management*, 25(2), 399-420.
- Arsenio, W. F., and Lemerise, E. A. (2004). Aggression and moral development: Integrating social information processing and moral domain models. *Child development*, 75(4), 987-1002.
- Aruna, S., Nandakishore, L. V., & Rajagopalan, S. P. (2012). Cloud Based Decision Support System for Diagnosis of Breast Cancer using Digital Mammograms. *International Journal of Computer Applications (IJCA)*, 1, 01-03.
- Bartol, C. R. (1991). Predictive validation of the MMPI for small-town police officers who fail. *Professional Psychology: Research and Practice*, 22(2), 127.
- Ben-Porath, Y. S., and Tellegen, A. (2011). *Manual for administration, scoring, and interpretation* (Vol. 45, No. 1). University of Minnesota Press. Pearson: Minneapolis, MN.
- Blas, LD, and Carraro, D.(2011). Relations between nonverbal intelligence and personality reports in late childhood. *Journal of Research in Personality*.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). MMPI-2: Manual for administration and scoring.
- Cohen, J. (1969). *Statistical power analysis for the behavioural sciences*. New York: Academic Press.
- Connor, K.M., & Davidson, J. R. T. (2003). *Development of a new resilience scale: the Connor-Davidson resilience Scale (CD-RISC)*. *Depression and Anxiety*, 18, 76-82.
- Correctional Service of Canada (2004). *Urinalysis statistics and drug seizures*. Ottawa: Security Division, Correctional Service of Canada.
- Correctional Service of Canada (2016). *Report of the Task Force on Security*. Ottawa: Security Division, Correctional Service of Canada.
- Cullen, F. T., Lutze, F. E., Link, B. G., and Wolfe, N. T. (1989). The correctional orientation of prison guards: Do officers support rehabilitation. *Fed. Probation*, 53, 33.
- Dahlstrom, W. G. (1992). Comparability of two-point high-point code patterns from original MMPI norms to MMPI-2 norms for the restandardization sample. *Journal of Personality Assessment*, 59(1), 153-164.
- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook: Clinical interpretation*, Vol. I, Rev.

- Davis, R. and Rostow, C. (2017). *M-PULSE - Matrix-Predictive Uniform Law Enforcement Selection Evaluation*, from <http://www.mhs.com/product.aspx?gr=safandprod=mpulseandid=resources> [Accessed 26 Mar. 2017].
- Department of Justice. (1995). Employee Equity Act. S.C. 1995, c.44 retrieved from <http://laws-lois.justice.gc.ca> (retrieved May 31st 2017).
- DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., and Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing scholarship*, 39(2), 155-164.
- Ehrig, R., Dierkes, T., Schäfer, S., Röblitz, S., Tronci, E., Mancini, T., ... & Massini, A. (2016, January). An integrative approach for model driven computation of treatments in reproductive medicine. In *BIOMAT 2015: Proceedings of the International Symposium on Mathematical and Computational Biology* (p. 67). World Scientific.
- Fischler, G. L. (1997). Psychological examination of peace officer applicants: Theory and practice. *Minnesota Police Chief*, 17, 35-37.
- Garg, A. X., Adhikari, N. K., McDonald, H., Rosas-Arellano, M. P., Devereaux, P. J., Beyene, J., ... & Haynes, R. B. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *Jama*, 293(10), 1223-1238.
- Graham, J. R. (1990). *MMPI-2: Assessing personality and psychopathology* (pp. 46-48). New York: Oxford University Press.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: a meta-analysis. *Psychological assessment*, 12(1), 19.
- Hathaway, S. R., McKinley, J. C., and MMPI Restandardization Committee. (1989). *MMPI-2: Minnesota Multiphasic Personality Inventory-2: manual for administration and scoring*. University of Minnesota Press.
- Hepburn, J. R., and Knepper, P. E. (1993). Correctional officers as human services workers: The effect on job satisfaction. *Justice Quarterly*, 10(2), 315-337.
- Holland, T. R., Heim, R. B., and Holt, N. (1976). Personality patterns among correctional officer applicants. *Journal of clinical psychology*, 32(4), 786-791.
- Huberty, C. J. (1975). Discriminant analysis. *Review of Educational Research*, 45(4), 543-598.
- Ingram, R. E., & Luxton, D. D. (2005). "Vulnerability-stress mode." In B. L. Hankin & J. R. Z. Abela (Eds.) *Development of Psychopathology: A vulnerability stress perspective* (pp.32-46). Thousand Oaks, CA: Sage Publications Inc.
- Kappeler, V. E., Sapp, A. D., & Carter, D. L. (1992). Police officer higher education, citizen complaints and departmental rule violations. *Am. J. Police*, 11, 37.
- Keen, P. G. (1987). Decision support systems: the next decade. *Decision Support Systems*, 3(3), 253-265.

- Keen, P. G., & Morton, M. S. S. (1978). *Decision support systems: an organizational perspective* (Vol. 35). Reading, MA: Addison-Wesley.
- Kitaeff, J. (2011). History of police psychology. *Handbook of Police Psychology*.—New York, London: Routledge, Taylor and Francis Group,—2011.—P, 1-60.
- Kong, G., Xu, D. L., Body, R., Yang, J. B., Mackway-Jones, K., & Carley, S. (2012). A belief rule-based decision support system for clinical risk assessment of cardiac chest pain. *European Journal of Operational Research*, 219(3), 564-573.
- Kornfeld, A. D. (1995). Police officer candidate mmpi-2 performance: Gender, ethnic, and normative factors. *Journal of Clinical Psychology*, 51(4), 536-540.
- Kuncel, N.R., Klieger, D. M., Connelley, B.S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology*, 98 (6), 1060-1072. Doi:10.1037/a0034156
- Linguistic Characteristics of Canadians. (2015, December 22). Retrieved March 21, 2017, from <http://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-314-x/98-314-x2011001-eng.cfm>
- Lombardo, L. X. (1981). *Guards imprisoned: Correctional officers at work* (pp. 162-63). New York: Elsevier.
- Maylor, E. A., and Rabbitt, P. (1993). Alcohol, reaction time and memory: A meta-analysis. *British Journal of Psychology*, 84(3), 301-317.
- McLachlan, G. (2004). *Discriminant analysis and statistical pattern recognition* (Vol. 544). John Wiley & Sons.
- MMPI History. (2015, September 04). Retrieved May 31, 2017, from <https://www.upress.umn.edu/test-division/bibliography/mmpi-history>
- Morgan, R. D., Van Haveren, R. A., and Pearson, C. A. (2002). Correctional officer burnout further analyses. *Criminal Justice and Behavior*, 29(2), 144-160.
- Ontario Provincial Police. (n.d.). Retrieved March 22, 2017, from <https://www.opp.ca/>
- Pearson (2015). *MMPI-2 Training Slides*, University of Minnesota Press
- Poulsen, J., & French, A. (2008). Discriminant function analysis
- Plourde, C., and Brochu, S. (2002). Drugs in prison: A break in the pathway. *Substance Use and Misuse*, 37, 47-63
- Principles Report for the validation and use of Personnel Selection Procedures (2003): retrieved from http://www.siop.org/_principles/principles.pdf 2016/12/14
- Richardson, J.T.E. (2011). Eta squared and partial eta squared as measurements of effect size in educational research. *Educational Research Review*, 6, 135-147.
- Schiele, B. C., Baker, A. B., and Hathaway, S. R. (1943). The Minnesota multiphasic personality inventory. *Journal-Lancet*.

- Sellbom, M., Fischler, G. L., and Ben-Porath, Y. S. (2007). Identifying MMPI-2 predictors of police integrity and misconduct. *Criminal Justice and Behavior*.
- Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1), 3-55.
- Steele, C. M., and Josephs, R. A. (1990). Alcohol myopia: Its prized and dangerous effects. *American Psychologist*, 45(8), 921.
- Steiner, B., and Wooldredge, J. (2015). Individual and environmental sources of work stress among prison officers. *Criminal Justice and Behavior*, 42(8), 800-818.
- Territo, L., and Vetter, H. J. (1981). Stress and police personnel. *Journal of Police Science and Administration*, 9(2), 195-208.
- Triplett, R., Mullings, J. L., and Scarborough, K. E. (1996). Work-related stress and coping among correctional officers: Implications from organizational literature. *Journal of Criminal Justice*, 24(4), 291-308.
- Waltman, S. H., Williams, A., & Christiansen, L. R. (2013). Comparing student clinician and licensed psychologist clinical judgment. *Training and Education in Professional Psychology*, 7(1), 33.
- Williams, K. M., Davis, R. D., & Rostow, C. D. (2011). Comparing the M-PULSE Inventory and MMPI-2: Empirical overlap and the prediction of liability outcomes in 7,161 law enforcement officers. In *annual meeting of the American Psychological Association*, Washington, DC.
- Weekes, J., Thomas, G., Graves, G. (2004). Substance Abuse in Corrections FAQs. *Canadian Centre on Substance Abuse (CCSA)*: retrieved from <http://www.ccsa.ca/Resource%20Library/ccsa-011058-2004.pdf>
- Zehr, H. (2015). *The little book of restorative justice: Revised and updated*. Skyhorse Publishing, Inc.